PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/54, 9/10 — A2

(11) International Publication Number: WO 92/09694

(43) International Publication Date: 11 June 1992 (11.06.92)

(21) International Application Number: PCT/CA91/00417

(22) International Filing Date: 29 November 1991 (29.11.91)

(30) Priority data:
620,098 30 November 1990 (30.11.90) US

(71) Applicant: HSC RESEARCH AND DEVELOPMENT LIMITED PARTNERSHIP [CA/CA]; 88 Elm Street, Toronto, Ontario M5G 1X8 (CA).

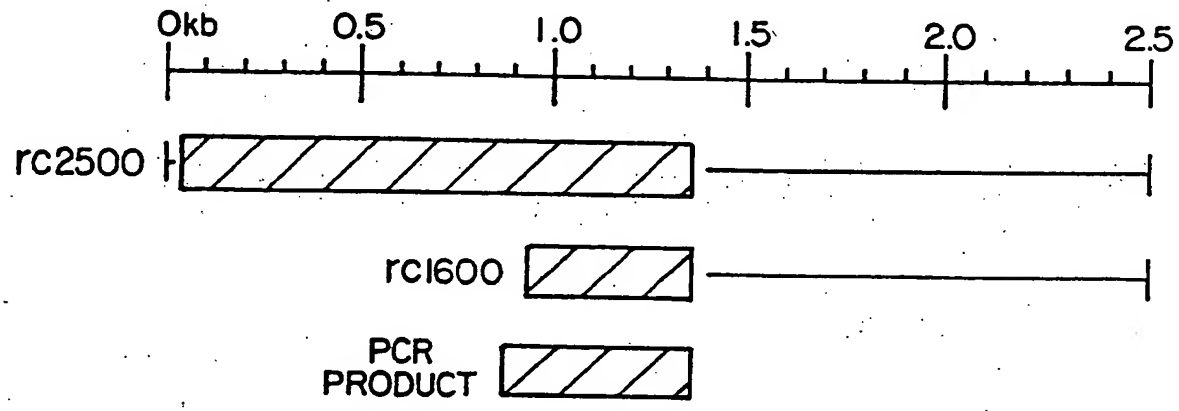
(72) Inventors: SCHACHTER, Harry ; 5 Menin Road, Toronto, Ontario M6C 3J1 (CA). SARKAR, Mohan ; 77 Elm Street, Apartment #405, Toronto, Ontario M5G 1H4 (CA).

(74) Agent: D'IORIO, Hélène; Gowling, Strathy & Henderson, Suite 2600, 160 Elgin Street, Ottawa, Ontario K1N 8S3 (CA).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, PL, RO, SD, SE, SE (European patent), SN (OAPI patent), SU+,TD (OAPI patent), TG (OAPI patent).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: CLONING OF UDP-N-ACETYLGLUCOSAMINE:α-3-D-MANNOSIDE β-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I



(57) Abstract

The genes encoding rabbit and human GnT I have been cloned.

+ See back of page

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

CLONING OF UDP-N-ACETYLGLUCOSAMINE:α-3-D-MANNOSIDE β-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to DNA sequences for the human and rabbit enzymes which control the conversion of high mannose to hybrid and complex N-glycans, UDP-N-acetylglucosamine:α-3-D-mannoside β-1,2-N-acetylglucosaminyltransferase I (GnT I), plasmids containing such DNA sequences, transformed cells containing such plasmids, and a method for converting high mannose glycoproteins to branched N-glycan glycoproteins.

Discussion of the Background

The biosynthesis of highly branched N- and O-glycans is important to many biological phenomena (Rademacher et al (1988) Ann. Rev. Biochem., vol. 57, 785-838). For example, baby hamster kidney cells transformed either by polyoma virus or by Rous sarcoma virus show a two-fold increase in one of the N-acetylglucosaminyltransferases (GlcNAc-transferase V) involved in the synthesis of highly branched complex N-glycans (Pierce et al (1986) J. Biol. Chem., vol. 261, 10772-10777; Yamashita et al (1985) J. Biol. Chem., vol. 260, 3963-3969). All N-glycans share the common core structure Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ-Asn. Complex N-glycans have "antennae" or branches attached to this core. The antennae are initiated by the action of at least five Golgi-localized membrane-bound GlcNAc-transferases designated GnT I, II, IV, V and VI (Schachter et al (1989) Methods Enzymol., vol. 179, 351-396) and may be further elongated by the addition of D-galactose, L-fucose and sialic acid residues. Complex N-glycans may be "bisected" by a GlcNAc residue attached in β1-4 linkage to the β-linked Man of the core due to the action of GlcNAc-transferase III (GnT III).

The conversion of high-mannose to complex and hybrid N-glycans is controlled by UDP-GlcNAc:α-3-D-mannoside β-1,2-N-

acetylglucosaminyltransferase I (GnT I, EC 2.4.1.101), which catalyzes the reaction:

UDP-GlcNAc + (Manα1-6[Manα1-3]Manα1-6) (Manα1-3)Manβ1-4R → (Manα1-6[Manα1-3]Manα1-6) (GlcNAcβ1-2Manα1-3)Manβ1-4R + UDP,

where R is GlcNAcβ1-4(+/-Fucα1-6)GlcNAc-Asn-X, and Asn-X may be an Asn residue which is part of the amino acid sequence of a protein.

The enzyme is specific for the Manα1-3Manβ1-4GlcNAc-arm of the core. The presence of a β2-linked GlcNAc residue at the non-reducing terminus of this arm is essential for subsequent action of several enzymes in the processing pathway (Schachter et al (1983) Can. J. Biochem. Cell Biol., vol. 61, 1049-1066; Schachter et al (1985) "Glycosyltransferases involved in the biosynthesis of protein-bound oligosaccharides of the asparagine-N-acetyl-D-glucosamine and serine(threonine)-N-acetyl-D-galactosamine types", in: A.N. Martonosi, ed. The Enzymes of Biological Membranes, New York, N.Y., Plenum Press, 227-277; Schachter, (1986) Biochem. Cell Biol., vol. 64, 163-181; Schachter (1988) Biochemie., vol. 70(11), 1701-1702), i.e., GnT II, III and IV require the prior action of GnT I, and GnT V and VI require the prior action of GnT II. GnT I has been reported in hen oviduct, Chinese hamster ovary cells, baby hamster kidney cells, bovine colostrum, pig trachea and mammalian liver (Schachter et al (1983) Can. J. Biochem. Cell Biol., vol. 61, 1049-1066; Schachter et al (1985) "Glycosyltransferases involved in the biosynthesis of protein-bound oligosaccharides of the asparagine-N-acetyl-D-glucosamine and serine(threonine)-N-acetyl-D-galactosamine types", in: A.N. Martonosi, ed. The Enzymes of Biological Membranes, New York, N.Y., Plenum Press, 227-277; Schachter et al (1980) "Mammalian glycosyltransferases: their role in the synthesis and function of complex carbohydrates and glycolipids", in:

Lennarz W.J., ed. Biochemistry of Glycoproteins and Proteoglycans, New York, N.Y., Plenum Press, 85-160; Brockhausen et al (1988) Biochem. Cell Biol., vol. 66, 1134-1151). The enzyme has been partially purified from bovine colostrum (Harpaz et al (1980) J. Biol. Chem., vol. 255, 4885-4893) and from pig liver and trachea (Oppenheimer et al (1981) J. Biol. Chem., vol. 256, 11477-11482), and to homogeneity from rabbit liver (Oppenheimer et al (1981) J. Biol. Chem., vol. 256, 799-804; Nishikawa et al (1988) J. Biol. Chem., vol. 263, 8270-8281).

Recently, the cloning of DNA encoding proteins and the expression of such cloned DNA to produce the proteins has become commercially important. For ease of culturing, it is preferred that the cloned DNA be expressed in a primitive host, such as a bacteria (e.g., E. coli), a yeast, or a fungus. However, such primitive hosts may not normally possess the enzymes required for the post-translation modification of proteins which occurs in the cells from which the DNA originated. Thus, although many primitive hosts possess the necessary enzymes to effect the post-translation modification of a protein to a high mannose derivative, such host do not contain the enzyme required to convert the high mannose derivative to a hybrid and branched glycan, GnT I.

As discussed in Bergh et al, "Glycosylation of Heterologously Expressed Proteins: Problems and Solutions", in Therapeutic Peptide and Proteins: Assessing the New Technologies, Marshak et al eds, Cold Spring Harbor Laboratory, Banbury Report 29, 1988, in prokaryotes, the resulting lack of glycosylation may have a variety of consequences, such as incorrect polypeptide chain-folding, precipitation and aggregation of the protein, proteolytic degradation or enhanced immunogenicity.

Yeast and vertebrate cells use the same $Glc_3Man_9GlcNAc_2$ lipid-linked precursor for cotranslational glycosylation of asparagine residues, both recognize the same Asn-X-ser/Thr sequences, and both remove the three glucose residues soon

after transfer. Thus, a mammalian glycoprotein expressed in yeast may contain the same carbohydrate chains as the native protein until after it leaves the endoplasmic reticulum. After entry into the Golgi, however, the later steps in oligosaccharide processing are very different in yeast (see Kukuruzinska et al, Ann. Rev. Biochem., vol. 56, p.915, 1987) and vertebrates, (see Hubbard and Ivatt Ann. Rev. Biochem., vol. 50, p.555, 1981; Kornfeld and Kornfeld Ann. Rev. Biochem., vol. 54, p.631, 1985). Processed Saccharomyces cerevisiae N-linked oligosaccharides contain two GlcNAc residues and from 9 to 50 or more mannose residues. On the other hand, mammalian oligosaccharides never have more than nine mannose residues and most commonly contain GlcNAc, galactose, and sialic acid attached to a $Man_3GlcNAc_2$ core.

Thus, heterologous expression in yeast of a mammalian glycoprotein intended for therapeutic use can present a number of potential glycosylation-related problems. For example, carbohydrate chains may be highly antigenic; in addition, they are recognized by Man/GlcNAc-specific receptors on cells of the mammalian reticuloendothelial system, resulting in rapid clearance of the glycoprotein from the circulation.

Thus, it is desirable to: (1) provide large amounts of GnT I for the further post translational modification of recombinantly produced proteins; and (2) provide a means for enabling primitive hosts to express GnT I.

However, as yet there are no methods available for obtaining large quantities of GnT I or enabling primitive hosts to express GnT I.

SUMMARY OF THE INVENTION

Accordingly, it is an object of the present invention to provide a method for producing large quantities of GnT I.

It is another object to provide a method for converting high mannose derivatives to hybrid and complex N-glycans.

It is another object to provide isolated DNA sequences which encode GnT I.

It is another object to provide plasmids which contain a DNA sequence which encodes GnT I.

It is another object to provide microorganisms which contain a heterologous sequence of DNA which encodes GnT I.

These and other objects, which will become apparent during the following detailed description, have been achieved by the inventors' isolation and cloning of DNA sequences encoding rabbit and human GnT I, preparation of plasmids containing such DNA sequences and transfection of microorganisms, with such plasmids.

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same become better understood by reference to the following detailed description when considered in connection with the accompanying drawings, wherein:

Figure 1 illustrates the amino acid sequence data for the eight peptides isolated from rabbit liver GnT I and nucleotide sequences of the six synthetic oligonucleotides prepared on the basis of the peptide sequences. The single letter code is used for amino acid sequence data; upper case letters indicate firm assignments and lower case letters indicate tentative assignments. The underlined sections of the peptide sequences indicate the regions used for the design of oligonucleotide probes. Probes 2, 3 and 6 were based on peptides 2, 3 and 6, respectively; S indicates "sense" and A indicates "antisense" directions;

Figure 2 illustrates a schematic representation of GnT I clones. PCR product, product obtained by PCR amplification of rabbit liver cDNA; rc 1600, 1.6 kb GnT I cDNA clone; rc2500, 3.0 kb GnT I cDNA clone. The shaded boxes represent the coding region. During subcloning, the 3.0 kb cDNA was reduced to 2.5 kb by a 0.5 kb deletion at the 5'-end;

Figure 3 illustrates the results of an agarose gel electrophoresis (1% agarose) of the products of the polymerase chain reaction (PCR) using rabbit liver cDNA as template and the following combinations of oligonucleotides as primers: 2S-3A; 2S-6A; 3S-2A; 3S-6A; 6S-2A; 6S-3A (Figure 1). Conditions of PCR are given in the Methods section. The gel was stained with ethidium bromide (0.5 μg/ml). Primer-dependent products were obtained with combinations 2S-6A (0.50 kb) and 3S-6A (0.45 kb). The arrow designates the 0.5 kb DNA marker; the remaining standards are at 1.0 kb, 1.6 kb, 2.0 kb and at 1.0 kb intervals thereafter;

Figure 4 illustrates the nucleotide sequence (lower case) of the 2.5 kb GnT I cDNA clone. The amino acid sequence in the coding region is shown in upper case letters. The positions of the eight peptide sequences obtained from proteolytic digests of GnT I (Figure 1) are underlined with a single solid line; the regions of these peptide sequences used for oligonucleotide probe synthesis (Figure 1) are additionally underlined with a discontinuous line. The putative transmembrane segment (bases 62-136) is underlined with a double line. The consensus polyadenylation signal AATAAA at position 2435 is underlined. Only the nucleotide sequence is numbered;

Figure 5 illustrates an autoradiogram of an SDS-polyacrylamide gel electrophoresis experiment showing <u>in vitro</u> transcription and translation of the rabbit cDNA. mRNA was generated from the 2.5 kb GnT I cDNA and was used as the template for <u>in vitro</u> translation using rabbit reticulocyte lysate and L-[^{35}S]-methionine (see Methods for details). Lane C, no plasmid in the incubation; lane 12, pGEM-7z containing the 2.5 kb GnT I cDNA with an insert between bases 56 and 57 which interrupts the reading frame; lane 16, pGEM-7z containing the 2.5 kb GnT I cDNA (pGEM-7z-rcgntl);

Figure 6 illustrates the nucleotide sequence for human genomic DNA encoding for GnT I;

Figure 7 illustrates the amino acid sequence for human GnT I; and

Figure 8 illustrates both the nucleotide sequence for human genomic DNA encoding for GnT I and the amino acid sequence of human GnT I.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Thus, one aspect of the present invention relates to isolated DNA sequences which encode rabbit GnT I. Specifically, such DNA sequences encode a protein having the sequence (starting from the N-terminal) of formula I shown below:

MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE
LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR
ARG PRO VAL PRO SER ARG LEU PRO SER ASP ASN ALA LEU ASP ASP
ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP
ALA GLU VAL GLU LEU GLU ARG GLN ARG GLY LEU LEU GLN GLN ILE
ARG GLU HIS HIS ALA LEU TRP SER GLN ARG TRP LYS VAL PRO THR
ALA ALA PRO PRO ALA GLN PRO HIS VAL PRO VAL THR PRO PRO PRO
ALA VAL ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL
ARG ARG CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU
LEU PHE PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR
ALA GLN VAL ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG
GLN PRO ASP LEU SER ASN ILE ALA VAL GLN PRO ASP HIS ARG LYS
PHE GLN GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU
GLY GLN ILE PHE HIS ASN PHE ASN TYR PRO ALA ALA VAL VAL VAL
GLU ASP ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE GLN
ALA THR TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL
SER ALA TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP SER SER
LYS PRO GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY
TRP LEU LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP
PRO LYS ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG
LYS GLY ARG ALA CYS VAL ARG PRO GLU ILE SER ARG THR MET THR
PHE GLY ARG LYS GLY VAL SER HIS GLY GLN PHE PHE ASP GLN HIS
LEU LYS PHE ILE LYS LEU ASN GLN GLN PHE VAL PRO PHE THR GLN
LEU ASP LEU SER TYR LEU GLN GLN GLU ALA TYR ASP ARG ASP PHE
LEU ALA ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL

ARG THR ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR
THR GLY ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL
MET ASP ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY
ILE VAL THR PHE LEU PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO
PRO GLN THR TRP ASP GLY TYR ASP PRO SER TRP THR

In another aspect, the present invention relates to DNA sequences which encode human GnT I. Such DNA sequences encode a protein having the sequence (starting from the N-terminus) of formula II shown below:

1: MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE
16: LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR
31: ARG PRO ALA PRO GLY ARG PRO PRO SER VAL SER ALA LEU ASP GLY
46: ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP
61: ALA GLU VAL GLU LEU GLU ARG ARG ARG GLY LEU LEU GLN GLN ILE
76: GLY ASP ALA LEU SER SER GLN ARG GLY ARG VAL PRO THR ALA ALA
91: PRO PRO ALA GLN PRO ARG VAL PRO VAL THR PRO ALA PRO ALA VAL
106: ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL ARG ARG
121: CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU LEU PHE
136: PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR ALA GLN
151: ALA ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG GLN PRO
166: ASP LEU SER SER ILE ALA VAL PRO PRO ASP HIS ARG LYS PHE GLN
181: GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU GLY GLN
196: VAL PHE ARG GLN PHE ARG PHE PRO ALA ALA VAL VAL VAL GLU ASP
211: ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE ARG ALA THR
226: TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL SER ALA
241: TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP ALA SER ARG PRO
256: GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY TRP LEU
271: LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP PRO LYS
286: ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG GLN GLY
301: ARG ALA CYS ILE ARG PRO GLU ILE SER ARG THR MET THR PHE GLY
316: ARG LYS GLY VAL THR HIS GLY GLN PHE PHE ASP GLN HIS LEU LYS
331: PHE ILE LYS LEU ASN GLN GLN PHE VAL HIS PHE THR GLN LEU ASP
346: LEU SER TYR LEU GLN ARG GLU ALA TYR ASP ARG ASP PHE LEU ALA
361: ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL ARG THR
376: ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR THR GLY
391: ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL MET ASP
406: ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY ILE VAL

421: THR PHE GLN PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO PRO PRO

436: THR TRP GLU GLY TYR ASP PRO SER TRP ASN

Exemplary of the DNA sequences encoding rabbit GnT I is the sequence (starting from the 5'-terminus) of formula III, shown below:

atg ctg aag aag cag tct gct ggg ctt gtg ctg tgg ggt gct atc
ctc ttt gtg gcc tgg aat gcc ctg ctg ctc ctc ttc ttc tgg aca
cgt cca gtg cct agc agg ctg ccg tca gac aat gct ctc gat gat
gac cct gcc agc ctc acc cgt gag gtg atc cgc tta gct cag gat
gcc gag gta gag ttg gaa cgt cag cgg gga ctg ttg cag cag att
agg gag cac cat gct ctt tgg agc cag cgg tgg aag gtg cct act
gca gcc cct cct gct cag ccg cat gtg cct gtg acc cca ccg cca
gct gtg atc ccc atc ctg gta att gcc tgt gac cgc agc acc gtc
cgc cgc tgt ttg gac aag cta ctg cat tat cgg cct tca gct gag
ctg ttc ccc atc att gtc agc cag gac tgt ggg cat gag gag aca
gcc cag gtc att gct tcc tat ggc agc gca gtc aca cac atc cgg
caa cct gac ctg agc aac att gct gtg cag ccc gac cac cgc aag
ttc cag ggc tac tac aag atc gca cgg cat tac cgc tgg gca ttg
ggc caa atc ttc cac aat ttc aac tac cca gca gct gtg gtg gtg
gag gat gat ctc gag gtg gca cca gac ttc ttt gag tac ttc cag
gcc act tac cca ctg ttg aaa gca gac ccc tcc ctc tgg tgt gtg
tct gcc tgg aat gac aat ggc aaa gaa cag atg gta gac tcg agt
aag cca gag tta ctc tac cgc aca gat ttc ttt cct ggc tta ggc
tgg tta ctg ttg gct gaa ctc tgg gct gaa ctg gag ccc aag tgg
ccc aaa gcc ttc tgg gat gac tgg atg cgc cgg cct gag cag cga
aag ggg agg gcc tgt gtg cgt cca gaa atc tca aga aca atg aca
ttt ggc cgg aag ggt gtg agc cat ggg cag ttc ttt gac cag cat
ctc aag ttc atc aag ctg aac cag cag ttt gta ccc ttc acc cag
ctg gac ctg tcg tac ctt cag cag gag gcc tat gac cgg gat ttc
ctt gct cgt gtt tat ggt gct ccc cag tta cag gtg gag aaa gtg
agg acc aat gac cgg aag gag cta gga gag gtg cgc gta cag tac
aca ggc agg gac agc ttc aag gct ttc gcc aag gcc ctg ggt gtc
atg gat gac ctc aaa tca ggt gta ccc agg gct gga tac cgg ggc
att gtc acc ttc tta ttc cgg ggc cgc cgt gtc cac ctg gcg ccc
cct cag act tgg gat ggc tat gat cct agt tgg act

The DNA sequence of formula III corresponds to the coding region of rabbit cDNA encoding GnT I. Another

example of a DNA sequence encoding rabbit GnT I is a larger section of cDNA encoding rabbit GnT I, which has the formula IV as shown below:

1 gaattccggc aagtcatacc tttgcctgcc ctcccctgtg ggggccagg
atg ctg aag aag cag tct gct ggg ctt gtg ctg tgg ggt gct atc
ctc ttt gtg gcc tgg aat gcc ctg ctg ctc ctc ttc ttc tgg aca
cgt cca gtg cct agc agg ctg ccg tca gac aat gct ctc gat gat
gac cct gcc agc ctc acc cgt gag gtg atc cgc tta gct cag gat
gcc gag gta gag ttg gaa cgt cag cgg gga ctg ttg cag cag att
agg gag cac cat gct ctt tgg agc cag cgg tgg aag gtg cct act
gca gcc cct cct gct cag ccg cat gtg cct gtg acc cca ccg cca
gct gtg atc ccc atc ctg gta att gcc tgt gac cgc agc acc gtc
cgc cgc tgt ttg gac aag cta ctg cat tat cgg cct tca gct gag
ctg ttc ccc atc att gtc agc cag gac tgt ggg cat gag gag aca
gcc cag gtc att gct tcc tat ggc agc gca gtc aca cac atc cgg
caa cct gac ctg agc aac att gct gtg cag ccc gac cac cgc aag
ttc cag ggc tac tac aag atc gca cgg cat tac cgc tgg gca ttg
ggc caa atc ttc cac aat ttc aac tac cca gca gct gtg gtg gtg
gaa gat gat ctc gag gtg gca cca gac ttc ttt gag tac ttc cag
gcc act tac cca ctg ttg aaa gca gac ccc tcc ctc tgg tgt gtg
tct gcc tgg aat gac aat ggc aaa gaa cag atg gta gac tcg agt
aag cca gag tta ctc tac cgc aca gat ttc ttt cct ggc tta ggc
tgg tta ctg ttg gct gaa ctc tgg gct gaa ctg gag ccc aag tgg
ccc aaa gcc ttc tgg gat gac tgg atg cgc cgg cct gag cag cga
aag ggg agg gcc tgt gtg cgt cca gaa atc tca aga aca atg aca
ttt ggc cgg aag ggt gtg agc cat ggg cag ttc ttt gac cag cat
ctc aag ttc atc aag ctg aac cag cag ttt gta ccc ttc acc cag
ctg gac ctg tcg tac ctt cag cag gag gcc tat gac cgg gat ttc
ctt gct cgt gtt tat ggt gct ccc cag tta cag gtg gag aaa gtg
agg acc aat gac cgg aag gag cta gga gag gtg cgc gta cag tac
aca ggc agg gac agc ttc aag gct ttc gcc aag gcc ctg ggt gtc
atg gat gac ctc aaa tca ggt gta ccc agg gct gga tac cgg ggc
att gtc acc ttc tta ttc cgg ggc cgc cgt gtc cac ctg gcg ccc
cct cag act tgg gat ggc tat gat cct agt tgg act
taacagctcc tgcctgtccc ttctgggctc cttccttgca atttcatgat ctaagatggg
accgtagtcc ctgggctgca ttgtcttttc tgtctttccc tcttgggtcc attttttttt
ttttcttttt tgagtggcat ttgaatacac agatgacaag gtgagggttc ttttgttaaa
ggagttagat cagggaaagc attctgctgt ctgttgggta tcaagcagca aaccactgtg
tgatagggga agaatgggct ttttggggcc agaaatatcc atgttctgag tttttctctt

aggtcatctg cagaggagtt ggcaacttta gctttcttaa ccaggccttt tctttctgac
ctgagagcca gggcatgaga cttcttgttc atgctccttt ttaccttccc ctaataaggg
tctgggctac aggagaagtg aacatattgt ggccagaata atactaacca gaggggcctc
attgtcagag tctaggtgca gttattgggt tgtcagagtt aatgccttct gttcttcttt
ccttattcct gacttctgtc agctcttctt tctttgcagc ctagcaattt ttggttctaa
gatgaaaaat gaagaggaaa agaaatattc gcacccagct attgggagaa aggtagtggg
aaaaaaactt cattgtacca cttcaaagag acactcttga cctcttcctt tctaaaaatt
agtcccctcc ctgttgcttc aggagaatgc tgtgctggtc agttctgtgt gatccttctt
ccctgagttt tatacacagg ctcctcccta aggctgtggc ttctggtggc cctcctgaca
taagttacag tggccaagac caggacaact ccggccatga gctaagtcct gcctaccttc
tccaaaacat tcccatgtcc tcacaggcta ggatgcagat gttggttgga gaggaatttg
tgtgtgtgtg tgtgtgtgtg tgtgttttct tgcctgacct cagtttcatg gatgaaaagt
ggaagctaca gaattatttt caaaaataaa ggctgaattg tctgaaaaaa aaaaaaaaaa
aaaaaaccgg aattc

The DNA sequence of formulae III and IV have been obtained by cloning the rabbit cDNA encoding GnT I, by the procedure which is described in detail in the Examples section.

Exemplary of the DNA sequences encoding human GnT I is the sequence (starting at the 5'-terminus) of formula V, shown below:

atgctgaa gaagcagtct gcagggcttg tgctgtgggg cgctatcctc tttgtggcct
961 ggaatgccct gctgctcctc ttcttctgga cgcgcccagc acctggcagg ccaccctcag
1021 tcagcgctct cgatggcgac cccgccagcc tcacccggga agtgattcgc ctggcccaag
1081 acgccgaggt ggagctggag cgcaggcgtg ggctgctgca gcagatcggg gatgccctgt
1141 cgagccagcg ggggagggtg cccaccgcgg cccctcccgc ccagccgcgt gtgcctgtga
1201 ccccgcgccc ggcggtgatt cccatcctgg tcatcgcctg tgaccgcagc actgttcggc
1261 gctgcctgga caagctgctg cattatcggc cctcggctga gctcttcccc atcatcgtta
1321 gccaggactg cgggcacgag gagacggccc aggccatcgc ctcctacggc agcgcggtca
1381 cgcacatccg gcagcccgac ctgagcagca ttgcggtgcc gccggaccac cgcaagttcc
1441 agggctacta caagatcgcg cgccactacc gctgggcgct gggccaggtc ttccggcagt
1501 ttcgcttccc cgcggccgtg gtggtggagg atgacctgga ggtggccccg gacttcttcg
1561 agtactttcg ggccacctat ccgctgctga aggccgaccc ctccctgtgg tgcgtctcgg
1621 cctggaatga caacggcaag gagcagatgg tggacgccag caggcctgag ctgctctacc
1681 gcaccgactt tttccctggc ctgggctggc tgctgttggc cgagctctgg gctgagctgg
1741 agcccaagtg gccaaaggcc ttctgggacg actggatgcg gcggccggag cagcggcagg
1801 ggcgggcctg catacgccct gagatctcaa gaacgatgac ctttggccgc aagggtgtga
1861 cgcacgggca gttctttgac cagcacctca agtttatcaa gctgaaccag cagtttgtgc
1921 acttcaccca gctggacctg tcttacctgc agcgggaggc ctatgaccga gatttcctcg
1981 cccgcgtcta cggtgctccc cagctgcagg tggagaaagt gaggaccaat gaccggaagg
2041 agctgggggа ggtgcgggtg cagtatacgg ggagggacag cttcaaggct ttcgccaagg
2101 ctctgggtgt tatggatgac cttaagtcgg gggttccgag agctggctac cggggtattg
2161 tcaccttcca gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct
2221 atgatcctag ctggaat

The DNA sequence of formula V corresponds to the coding region of human genomic DNA encoding GnT I. Another example of a DNA sequence encoding human GnT I is a larger section

of human genomic DNA encoding GnT I, which has the formula VI, shown below:

```
   1 aagttttgaa tgtttaagtt tatttaagtt tatttctaaa tattttctca tttctctggc
  61 ttttgtaagt agggttttct catccatgtt ttcttctcat gagttatttg tggatatgaa
 121 ggctatccat tagtatatgt tgatttttat attacacttc cttgctcagt tcattattga
 181 ttctttttga gttttccagg catattctca caagtaaaga taatagaaat agtttgcttc
 241 ctttccactt ctgctttgaa tttttttttc ttggttcatt tgcattggct gcttcctcca
 301 gcaaaatgtt aaataaccct ggagatgatg ggcaacttcg ttttgctcct gacattcgtg
 361 gggtgcctct ggtgcttccc tgttggtaag gggttaactg tagccctgag gtgggacatt
 421 tgattttaaa aatcagtcat cttggggcgc ttaggttaga ggaatggtag gcagatgctg
 481 tcactccttg ccccтcccct cctccttccc acctggaggg gaaatgaaat ctgacaggta
 541 gaaagagggg agttggggtt ctttttctct ctcccтccac cagcatcact ctctgcctct
 601 ccctcaaaaa tacgttcctg ggtcaggata tatgttgact ccctagagag ctctggagtc
 661 aacctcctgg ccttcctcca ccctcactct tggccttttc ctgcccccat ttcctctacc
 721 tgtggggcat ggagccacga gcctttgtgt gacggtttgc tttctctctc ctgtctttag
 781 gtgcatggct gcctcctaat cccatagtcc agaggaggca tccctaggac tgcgggcaag
 841 ggagccgcaa gcccagggca gccttgaacc gtcccctggc ctgccctccg gtgggggcca
 901 ggatgctgaa gaagcagtct gcagggcttg tgctgtgggg cgctatcctc tttgtggcct
 961 ggaatgccct gctgctcctc ttcttctgga cgcgcccagc acctggcagg ccaccctcag
1021 tcagcgctct cgatggcgac cccgccagcc tcacccggga agtgattcgc ctggcccaag
1081 acgccgaggt ggagctggag cgcaggcgtg ggctgctgca gcagatcggg gatgccctgt
1141 cgagccagcg ggggagggtg cccaccgcgg ccctctcccgc ccagccgcgt gtgcctgtga
1201 cccccgcgcc ggcggtgatt cccatcctgg tcatcgcctg tgaccgcagc actgttcggc
1261 gctgcctgga caagctgctg cattatcggc cctcggctga gctcttcccc atcatcgtta
1321 gccaggactg cgggcacgag gagacggccc aggccatcgc ctcctacggc agcgcggtca
1381 cgcacatccg gcagcccgac ctgagcagca ttgcggtgcc gccggaccac cgcaagttcc
1441 agggctacta caagatcgcg cgccactacc gctgggcgct gggccaggtc ttccggcagt
1501 ttcgcttccc cgcggccgtg gtggtggagg atgacctgga ggtggccccg gacttcttcg
1561 agtactttcg ggccacctat ccgctgctga aggccgaccc ctccctgtgg tgcgtctcgg
1621 cctggaatga caacggcaag gagcagatgg tggacgccag caggcctgag ctgctctacc
1681 gcaccgactt tttccctggc ctgggctggc tgctgttggc cgagctctgg gctgagctgg
1741 agcccaagtg gccaaaggcc ttctgggacg actggatgcg gcggccggag cagcggcagg
1801 ggcgggcctg catacgccct gagatctcaa gaacgatgac ctttggccgc aagggtgtga
1861 cgcacgggca gttctttgac cagcacctca agtttatcaa gctgaaccag cagtttgtgc
1921 acttcaccca gctggacctg tcttacctgc agcgggaggc ctatgaccga gatttcctcg
1981 cccgcgtcta cggtgctccc cagctgcagg tggagaaagt gaggaccaat gaccggaagg
2041 agctgggggа ggtgcgggtg cagtatacgg ggagggacag cttcaaggct ttcgccaagg
2101 ctctgggtgt tatggatgac cttaagtcgg gggttccgag agctggctac cggggtattg
2161 tcaccttcca gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct
2221 atgatcctag ctggaattag cacctgcctg tccttcctgg gccccttctt gccacatcat
2281 gagctgaggt gaccacagtc cccaggctgc atcggcctgc ctgtgtttcc ctcttaggtg
2341 catttatctt tttgattttt ccgagtggca tttaagtgca caaatgataa caagaggatt
2401 attctcccgt tctcaaggga gtcagatcag gggaactatt ctagggtatg ttgcggggta
2461 ttaagcagga aaacactgtg tggtgggggg cactgggctt gttggggcca caaatgtcca
2521 cgtcctgagc tttctcctgg agcatgtgca gagagtttgg caacgttcgc tctcttgacc
2581 agacccccttc tccctgactg gctcttccag ccaggcacga gccctccttc tatacctgct
2641 ccccttccca gtggggactg agttatggga gaaggggaca tatttgtggc caaaatgata
2701 ctaaccaaag gggcttcctt gtcagggcct ggtggagttg gtgggtcatc ggggctcact
2761 gcctcctgcc cttctctcct gtctgacccc cacttagccc ttctctcctt gcagcctagc
2821 agtttatagt tctgagatgg aaagttgaag ggggcaagca agacctctcc tcagcccatg
2881 cccagctgtc aggagagagg tgcagggagg aaggccttgt gctgggacaa cctctctctt
2941 gccttacctt cagagaggac tatgccctga ccctcccttt ctgaaaatca gtgccctctc
3001 tgttgctcta ggaggctcct gctggcttgg tagaagacag aattcgatct gcctgtccct
3061 ttttcccctg gggtttgaca cacaggctcc tctcagcatg aggtggagca gtgaccaggt
```

3121 ggagcagtga ccaggacgcc tctggcccag tgctgcccag cctccccgcc cgctcccagg
3181 cgccccatgt cctcacaggc caggacgcca tggcggccgg gagcatgcga

The DNA sequences of formulae V and VI have been obtained by cloning human genomic DNA encoding GnT I, by the procedure which is described in detail in the Examples section.

Of course, it is to be understood that the present DNA sequences also include those which may not exactly match the sequences of formulae III-VI, but rather contain a small number of nucleotide substitutions, deletions, and/or additions. Further, the present DNA sequences also include those which encode for amino acid sequences which may not exactly match the sequences of formulae I and II, but rather contain a small number of amino acid residue substitutions, deletions, and/or additions, provided that the protein encoded by the DNA sequence exhibits GnT I activity.

In another embodiment, the present invention relates to plasmids which contain a DNA sequence encoding rabbit or human GnT I. Such plasmids may be prepared by conventional techniques and include plasmids formed by inserting one of the present DNA sequences into any suitable plasmid. Specific examples of the present plasmids include pGEM-7z-rcgnt1, in which a 2.5 kb sequence of rabbit cDNA encoding for GnT I (Figure 2) has been inserted into pGEM-7z; pGEX-2t-rcgnt1, in which a 2.5 kb sequence of rabbit cDNA encoding GnT I bas been inserted into pGEX-2t; and pGEM-5z-hggnti, in which a 4 kb sequence of human genomic DNA encoding GnT I has been inserted into pGEM-5z. The preparation of the plasmids pGEM-7z-rcgnt1, pGEX-2t-rcgnt1, and pGEM-5z-hggnt1 is described in detail in the Examples section, and all three of these plasmids have been deposited under the provisions of the Budapest Treaty with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852, USA on November 30, 1990 (Accession numbers not yet known).

In another embodiment, the present invention relates to transformed microorganisms which contain a heterologous

sequence of DNA encoding rabbit or human GnT I. Examples of suitable host cells including: bacteria, such as E. coli, Brevibacteria, and Coryneforms; fungus, such as Trichoderma reesei, Aspergillus niger, and Aspergillus awamori; yeast, such as Saccharomyces cerevisiae, Candida albicans, Candida utilis, Candida parapsilosis, Schizosaccharomyces pombe, Bandeiraea simplicifolia, Kluyveromyces lactis, Saccharomyces kluyveri, Hansenula, Saccharomycodes and Pichia; and vertebrate cells such as Chinese hamster ovary cells and COS cells. The transformed cells may be prepared by transfecting the cells with any of the present plasmids by conventional methods.

Another aspect of the present invention relates to methods for the production of GnT I. In a first embodiment, the present method comprises cell-free or in vitro expression of one of the present DNA sequences to obtain GnT I. For example, in vitro transcription and translation of one of the present plasmids using a system such as described in Methods in Molecular Biology, Nucleic Acids, Walker, ed., Humana Press, Clifton, NJ, pp 145-155 (1984) yields GnT I.

In another embodiment, the present method comprises culturing a microorganism which contains a heterologous DNA sequence which corresponds to one of the present DNA sequences. Although the culturing conditions, such as time, medium, temperature, light, and agitation, will depend on the identity of the host microorganism and the yield of GnT I desired, these conditions are readily determined by those skilled in the art.

In a further aspect, the present invention relates to a method for converting a glycoprotein which is in the high mannose form to a glycoprotein which is in the form of a hybrid or complex N-glycan. In a first embodiment, the present method may be carried out by reacting, in vitro, a glycoprotein which is in the high mannose form with mannosidases followed by UDP-GlcNAc in the presence of GnT I.

In another embodiment, the present method may comprise culturing a cell which produces a glycoprotein in high mannose form and which also contains a heterologous sequence of DNA encoding human or rabbit GnT I. For example, transfection of cell, which normally produces a glycoprotein in a mannose form, with one of the present plasmids may be used to form a cell which produces the protein (produced in high mannose form before transfection) as a hybrid or complex N-glycan. Preferably, the glycoprotein, which is produced in the high mannose form prior to transfection with the present DNA, is also produced by the host cell as a result of transformation. In other words, the DNA encoding the glycoprotein is also heterologous with respect to the host cell.

Examples of such glycoproteins are described in Tanner et al, Biochimica et Biophysica Acta; vol. 906, pp. 81-99 (1987); and Kukurazinska et al, Ann. Rev. Biochem., vol. 56, pp 915-944 (1987) and include SUC 2, CSF, c-IgM μ-chain, c-IgM chain, c-amylase, c-HBsAg, c-hemagglutinin, c-a_1 antitrypsin, c-preα_1, antitrypsin, c-glycoamylase, c-VSV gp, c-sindbis virus E1 yp, c-sindbis virus E2 gp, c-killerprotoxin (type I), c-phascolin α and β, hepatitis B virus surface antigen, interferon-gamma, tissue plasminogen activator, monoclonal anti-bodies, chicken ovalbumin-like proteins, interleukin-2, and proteins from vesicular stomatitis, influenza, and Semliki Forest viruses.

As noted above, branched glycans on membrane glycoproteins have been implicated in a variety of biological phenomena, e.g. tumor progression and metastasis, embryogenesis, cell differentiation, cell-cell and receptor-ligand interactions, viral and bacterial infectivity, fertilization and the control of the immune system (Rademacher et al (1988) Ann. Rev. Biochem., vol. 57, 785-838; Pierce et al (1986) J. Biol. Chem., vol. 261, 10772-10777; Yamashita et al (1985) J. Biol. Chem., vol. 260, 3963-3969; Schachter (1986) Biochem. Cell Biol., vol. 64, 163-181; West (1986) Mol. Cell. Biochem., vol. 72,

3-20; Narasimhan et al (1988) J. Biol. Chem., vol. 263, 1273-1281; Dennis et al (1987) Science, vol. 236, 582-585). GnT I catalyzes an essential first step in the conversion of high mannose to branched hybrid and complex N-glycans (Schachter (1986) Biochem. Cell Biol., vol. 64, 163-181; Brockhausen et al (1988) Biochem. Cell Biol., vol. 66, 1134-1151). In vitro transcription/translation of the 2.5 kb cDNA reported in this paper results in GnT I activity demonstrating the cloning of the gene for the catalytic domain of this important control enzyme.

At least seven glycosyltransferases involved in the synthesis of N- and O-glycans have been cloned to date, i.e., UDP-Gal:GlcNAc-R β1,4-Gal-transferase (Appert et al (1986) Biochem. Biophys. Res. Commons., vol. 139, 163-168; D'Agostaro et al (1989) Eur. J. Biochem., vol. 183, 211-217; Masri et al (1988) Biochem. Biophys. Res. Commun., vol. 157, 657-663; Narimatsu et al (1986) Proc. Nat. Acad. Sci. USA, vol. 83, 4720-4724; Shaper et al (1986) Proc. Nat. Acad. Sci. USA, vol. 83, 1573-1577; Shaper et al (1988) J. Biol. Chem., vol. 263, 10420-10428; Nakazawa et al (1988) J. Biochem. (Tokyo), vol. 104, 165-168). UDP-Gal:Gal-R α1,3-Gal-transferase (Joziasse et al (1989) J. Biol. Chem., vol. 264, 14290-14297; Larsen et al (1989) Proc. Natl. Acad. Sci. USA, vol. 86, 8227-8231; Larsen et al (1990) J. Biol. Chem., vol. 265, 7055-7061; Smith et al (1990) J. Biol. Chem., vol. 265, 6225-6234), CMP-sialic acid:Gal-R α2,6-sialyltransferase (Weinstein et al (1987) J. Biol. Chem., vol. 262, 17735-17743), CMP-sialic acid:Gal-R α2,3-sialyltransferase (Paulson et al (1990) FASEB J., vol. 4, A1862), GDP-Fuc:Galβ1,4(3)GlcNAc-R (Fuc to GlcNAc)α1,3(4)-Fuc-transferase (Gersten et al (1990) FASEB J., vol. 4, A1930; Kukowska-Latallo (1990) FASEB J., vol. 4, A1930), GDP-Fuc:Gal-R α1,2-Fuc-transferase (Rajan et al (1989) J. Biol. Chem., vol. 264(19), 11158-11167; Ernst et al (1989) J. Biol. Chem., vol. 264(6), 3436-3447) and UDP-GalNAc:Fucα1,2Gal-R (GalNAc to Gal) α1,3-GalNAc-transferase (Yamamoto et al (1990) J. Biol.

Chem., vol. 265, 1146-1151). These transferases all place sugars in terminal or subterminal positions; three of them (β1,4-Gal-, α2,6-sialyl-, and α1,3-GalNAc-transferases) have been localized to the trans-Golgi cisternae and trans-Golgi network, at least in some tissues (Roth et al (1982) J. Cell Biol., vol. 92, 223-229; Roth (1984) J. Cell Biol., vol. 98, 399-406; Roth (1987) Biochem. Biophys. Acta., vol. 906, 405-436; Roth et al (1988) Eur. J. Cell Biol., vol. 46, 105-112; Duncan et al (1988) J. Cell Biol., vol. 106, 617-628; Lee et al (1989) J. Biol. Chem., vol. 264, 13848-13855; Tooze et al (1988) J. Cell Biol., vol. 106, 1475-1487; Berger et al (1985) Proc. Nat. Acad. Sci. USA, vol. 82, 4736-4739; Taatjes et al (1988) J. Biol. Chem., vol. 263, 6302-6309). Human α1,3-GalNAc-transferase and a human pseudogene showing homology to murine α1,3-Gal-transferase share 55% homology (Laresen et al (1990) J. Biol. Chem., vol. 265, 7055-7061). CMP-sialic acid:Gal-R α2,6- and α2,3-sialyltransferases exhibit 50% identity and 80% conservation over a 50 amino acid stretch (Paulson et al (1990) FASEB J., vol. 4, A1862). The remaining transferases share no significant sequence similarities but have very similar domain structures, i.e., a short amino-terminal cytoplasmic tail, a 16-20 amino acid transmembrane segment (non-cleavable signal-anchor domain), a "stem" or "neck" region of undetermined length, and a long carboxyterminal catalytic domain which is in the Golgi lumen (Paulson et al (1989) J. Biol. Chem., vol. 264, 17615-17618).

The presence of a "neck" region is based on the finding that the α2,6-sialyltransferase (Weinstein et al (1987) J. Biol. Chem. vol. 262, 17735-17743; Lammers et al (1988) Biochem. J., vol. 256, 623-631) and the β1,4-Gal-transferase (D'Agostaro et al (1989) Eur. J. Biochem., vol. 183, 211-217) can be cut by proteases to release a smaller catalytically active protein lacking the trans-membrane domain. The exact length of this "neck" region cannot be stated with accuracy since it is not known how much of the amino-terminal sequence can be removed without loss of

catalytic activity. It has been shown that rabbit liver GnT I (Nishikawa et al (1988) J. Biol. Chem., vol. 263, 8270-8281) and rat liver UDP-GlcNAc:α-6-D-mannoside β-1,2-N-acetylglucosaminyltransferase II (GnT II) (Bendiak et al (1987) J. Biol. Chem., vol. 262, 5784-5790; Bendiak et al (1987) J. Biol. Chem., vol. 262, 5775-5783) exist in two forms, a large amount of presumably membrane-bound material which does not adhere to columns and a small amount of material which can be purified. In the case of GnT I, it is now clear from the sequence analysis that the 45 kDa form of the catalytically active protein previously purified has been derived from the membrane-bound precursor by proteolytic cleavage at about base position 215 in the "neck" region (Figure 4). The N-terminal blockage of this 45 kDa protein must therefore be due to chemical modification during GnT I purification. The hydrophobic trans-membrane region can form an α-helix with a hydrophobic surface capable of interacting with the membrane or with other hydrophobic proteins within the membrane. This strong hydrophobic interaction may explain why it is so difficult to purify glycosyltransferase preparations with intact trans-membrane domains.

Rabbit GnT I, human, mouse and bovine UDP-Gal:GlcNAc-R β1,4-Gal-transferases and human UDP-GalNAc:Fucα1,2Gal-R (GalNAc to Gal) α1,3-GalNAc-transferase have an abnormally high number of Pro residues between the transmembrane domain and the catalytic domain, e.g., there are 13 Pro residues in GnT I between the transmembrane domain and base position 376 (Figure 4); 9 of these Pro residues occur in a short stretch of 21 amino acids (bases 314-376, Figure 4). This Pro-rich "neck" may play a role in positioning the catalytic domain in the lumen of the Golgi to enable glycosylation of glycoproteins moving along the Golgi lumen.

The domain structure of GnT I appears to be similar to that of the previously cloned glycosyltransferases. However, GnT I differs from these transferases in being a medial-Golgi enzyme, at least in some tissues (Dunphy et al

(1985) Cell, vol. 40, 463-472; Kornfeld et al (1985) Ann. Rev. Biochem., vol. 54, 631-664). Although no medial-Golgi glycosyltransferase has been cloned to date, rat liver α-mannosidase II (also a medial-Golgi enzyme) has been partially cloned (Moremen (1989) Proc. Natl. Acad. Sci. USA, vol. 86(14), 5276-5280). Comparison with GnT I reveals a 16-amino acid sequence in GnT I (LHYRPSAELFPIIVSQ, bases 431-478, Figure 4) which shows a high similarity score to amino acid residues 403-418 in α-mannosidase II (LQYRNYEQLFSYMNSQ). Paulson's group (Paulson et al (1989) J. Biol. Chem., vol. 264, 17615-17618; Colley et al (1989) J. Biol. Chem., vol. 264, 17619-17622) has suggested that the trans-Golgi retention signal lies in the amino-terminal 57 amino acids of the α2,6-sialyltransferase molecule. The 16-amino acid "consensus" sequence present in GnT I and α-mannosidase II may be the equivalent medial-Golgi retention signal. Joziasse et al (1989) J. Biol. Chem., vol. 264, 14290-14297, have suggested that a column hexapeptide sequence K(R)DKKND(E) may serve as a UDP-Gal binding site in the β1,4-Gal- and α1,3-Gal-transferases; this sequence is not present in GnT I.

Sequence data indicate that the carboxy-terminal half of human GnT I shows 87% nucleotide sequence similarity and 90% amino acid sequence similarity to the carboxy-terminal half of rabbit liver GnT I. Strong homology between species has also been observed for bovine, murine and human UDP-Gal:GlcNAc-R β1,4-Gal-transferase (Appert et al (1986) Biochem. Biophys. Res. Commun., vol. 139, 163-168; D'Agostaro et al (1989) Eur. J. Biochem., vol 183, 211-217; Masri et al (1988) Biochem. Biophys. Res. Commun., vol. 157, 657-663; Narimatsu et al (1986) Proc. Nat. Acad. Sci. USA, vol. 83, 4720-4724; Shaper et al (1986) Proc. Nat. Acad. Sci. USA, vol. 83, 1573-1577; Shaper et al (1988) J. Biol. Chem., vol. 263, 10420-10428; Nakazawa et al (1988) J. Biochem. (Tokyo), vol. 104, 165-168) bovine and murine UDP-Gal:Gal-R α1,3-Gal-transferase (Joziasse et al (1989) J. Biol. Chem., vol. 264, 14290-14297; Larsen et al (1989)

Proc. Natl. Acad. Sci. USA, vol. 86, 8227-8231), murine and human GDP-Fuc:Galβ1,4(3)GlcNAc-R (Fuc to GlcNAc) α1,3(4)-Fuc-transferase (Gersten et al (1990) FASEB J., vol. 4, A1930; Kukowska-Latallo et al (1990) FASEB J., vol. 4, A1930), and human and rat CMP-sialic acid:Gal-R α2,6-sialyltransferase (Lance et al (1989) Biochem. Biophys. Res. Commun., vol. 164, 225-232).

It has been reported (Kumar et al (1990) Mol. Cell Biol., vol. 9, 5713-5717; Ripka et al (1989) Biochem. Biophys. Res. Commun. vol. 159(2), 554-560; Ripka et al (1990) J. Cellular Biochem., vol. 42, 117-122) that transformation of Lec I Chinese hamster ovary (CHO) cell mutants (which lack GnT I) with a crude preparation of total human genomic DNA results in transfectants expressing GnT I enzyme activity; this approach should allow cloning of the human GnT I gene by the gene transfer and expression screening method recently used to clone several glycosyltransferases (Larsen et al (1989) Proc. Natl. Acad. Sci. USA, vol. 86, 8227-8231; Larsen et al (1990) J. Biol. Chem., vol. 265, 7055-7061; Smith et al (1990) J. Biol. Chem., vol. 265, 6225-6234; Gersten (1990) FASEB J., vol. 4, A1930; Kukowska-Latallo et al (1990) FASEB J., vol. 4, A1930; Rajan et al (1989) J. Biol. Chem., vol. 264(19), 11158-11167; Ernst et al (1989) J. Biol. Chem., vol. 264(6), 3436-3447).

Other features of the invention will become apparent in the course of the following descriptions of exemplary embodiments which are given for illustration of the invention and are not intended to be limiting thereof.

EXAMPLES

I. Rabbit:

Preparation of Peptides. Rabbit liver GnT I was purified as previously described (Nishikawa et al (1988) J. Biol. Chem., vol. 263, 8270-8281). Glycerol, Triton X-100 and salts were removed from the purified enzyme (approximately 15 μg) by "inverse-gradient" reversed-phase

high performance liquid chromatography (RP-HPLC) (Simpson et al (1987) Eur. J. Biochem., vol. 165, 21-29). The enzyme solution (100 µl) was diluted to 1.2 ml with n-propanol in a sample-loading syringe, thoroughly mixed, and loaded at 1 ml/min on a VeloSep C_8 cartridge (3-µm particle size, 30 x 2.1 mm i.d.; Applied Biosystems, Foster City, CA, USA) previously equilibrated in 100% n-propanol at 40°C. GnT I was retained on the reversed-phase column under these conditions whereas glycerol, Triton X-100 and salts were washed through the column with 100% n-propanol. GnT I was eluted at 0.1 ml/min as a sharp peak by a linear gradient (5%/min) of decreasing n-propanol concentration (100% to 50%) generated with 100% n-propanol and 50% n-propanol/50% water containing 0.4% (v/v) trifluoroacetic acid at 40°C. GnT I-containing fractions from the inverse gradient RP-HPLC were pooled, adjusted to 0.02% (w/v) with respect to Tween 20 (Pierce Chemical Co., Rockford, IL, USA), concentrated to 100 µl in a 1.5-ml polypropylene tube using a centrifugal vacuum concentrator to reduce the n-propanol concentration, and diluted to 1.5 ml with 5% (v/v) formic acid containing 0.02% Tween 20.

Edman degradation of purified GnT I (~ 200 pmol) yielded no N-terminal sequence indicating N-terminal blockage; proteolysis of GnT I was therefore undertaken. GnT I was digested with pepsin (Sigma) at an enzyme/substrate mass ratio of 1:20 for 1 h at 37°C and the digest was fractionated by RP-HPLC on a short microbore column (30 x 2.1 mm i.d.) employing a low pH (trifluoroacetic acid, pH 2.1) mobile phase and a gradient of acetonitrile to yield peptides 5 and 6 (Figure 1). Core GnT I remaining after pepsin digestion was reduced with dithiothreitol and alkylated with iodoacetic acid (Simpson et al (1988) Eur. J. Biochem., vol. 176, 187-197) to give core S-carboxymethylated(SCM)-GnT I which was purified by RP-HPLC (Simpson et al (1988) Eur. J. Biochem., vol. 176, 187-197; Simpson et al (1989) Anal. Biochem., vol. 177, 221-236). Pepsin-treated core SCM-GnT I (about 10 µg in

1 ml 1% ammonium bicarbonate, 1mM $CaCl_2$, 0.02% Tween 20) was digested with trypsin (Worthington) at an enzyme/substrate mass ratio of 1:20 for 16 h at 37°C. RP-HPLC of the digest showed that trypsin resulted in little further digestion of the pepsin-treated material. Sequence analysis of a portion of this material resulted in 33 amino acid assignments (peptide 1, Figure 1). Pepsin and trypsin-treated core SCM-GnT I (about 8 μg in 1 ml 1% ammonium bicarbonate-0.02% Tween 20) was digested with thermolysin (Sigma) at an enzyme/substrate mass ratio of 1:20 for 2 h at 50°C and the digest was fractionated by RP-HPLC to yield peptides 2, 3, 4, 7 and 8 (Figure 1). Core GnT I was extremely resistant to proteolysis even after reduction and alkylation indicating that the molecule is probably very compact.

HPLC. RP-HPLC was carried out on a Hewlett-Packard liquid chromatograph (model 1090A) fitted with a diode array detector (model 1040A) (Simpson et al (1988) Eur. J. Biochem., vol. 176, 187-197). A Brownlee RP-300 column (30-nm pore size, 7-μm diameter dimethyloctylsilica particles packed into a stainless steel cartridge, 30 x 2.1 mm i.d.; Brownlee Laboratories, Santa Clara, CA, USA) was used for all peptide separations.

Amino Acid Sequence Analysis. Automated amino acid sequence analysis of GnT I and derived peptides was performed with Applied Biosystems sequencers (models 470A and 477A) equipped with on-line phenylthiohydantoin (PTH) amino acid analyzers (model 120A). Polybrene (Klapper et al (1978) Anal. Biochem., vol. 85, 126-131) was used as a carrier.

Oligonucleotides and cDNA Synthesis. Oligonucleotides were synthesized on a Pharmacia automated oligonucleotide synthesizer at the Hospital for Sick Children-Pharmacia Biotechnology Service Centre. Total RNA was prepared from rabbit liver by the method of Chirgwin et al (Chirgwin et al (1979) Biochemistry, vol. 18, 5294-5299; Ausubel et al (1990) Current Protocols in Molecular Biology, Media, PA:Greene Publishing Associates and John Wiley and Sons).

Poly(A)+RNA was prepared by oligo(dt) chromatography (Aviv et al (1972) Proc. Natl. Acad. Sci, USA, vol. 69, 1408-1412) using the mRNA Purification Kit supplied by Pharmacia. Single-stranded cDNA synthesis was performed using the RiboClone cDNA Synthesis System (Promega) with the following modifications. Total rabbit liver RNA (20 μg) in a volume of 5.5 μl was heated at 65°C for 3 min followed by cooling on ice for 5 min. The following reagents were added to a final volume of 50 μl:50 mM Tris-HCl, pH 8.3; 0.15 M KCl; 10 mM $MgCl_2$; 2 mM dithiothreitol (DTT); each dNTP at 0.4 mM; 40 units of RNasin (Promega); 2 mM sodium pyrophosphate; a mixture of the three anti-sense oligonucleotide primers 2A, 3A and 6A (Figure 1) at concentrations of 50 nM each; 20 units of AMV reverse transcriptase and 15 units of murine leukemia virus reverse transcriptase. Incubation was at 42°C for 2 hr. The reaction mixture was treated with NaOH (0.25 N final concentration) for 5 min at room temperature to destroy RNA. The solution was then heated at 65°C for 1 min followed by cooling on ice for 5 min and neutralized with HCl (0.25 N final concentration). This cDNA preparation was used directly in the PCR reaction.

Amplification of cDNA. PCR was carried out in a total volume of 0.1 ml containing 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 0.01% gelatin, each of the four dNTP at 0.2 mM, 0.5 μM of each oligonucleotide in six paired combinations of oligonucleotide primers (2S-3A, 2S-6A, 3S-2A, 3S-6A, 6S-2A, 6S-3A, Figure 1), 10 μl of RNA-free rabbit liver cDNA (see above), 2.5 units of Thermus aquaticus (Taq) polymerase (Perkin-Elmer/Cetus) and 0.1 ml of mineral oil. The samples were placed in an automated heating/cooling block (DNA Thermal Cycler, Perkin-Elmer) programmed for a temperature-step cycle of 94°C (0.5 min), 50°C (1 min) and 72°C (2 min) for a total of 40 cycles followed by a 10-minute extension at 72°C after the final cycle. DNA from the PCR reactions was purified with GeneClean (Bio 101, Inc.) and analyzed by electrophoresis in a 1% agarose gel containing ethidium bromide (0.5 μg/ml).

Two PCR products (0.45 and 0.50 kb) were detected and were purified from a 1% agarose gel by GeneClean. The DNA ends were filled in with T4 DNA polymerase (Moremen (1989) Proc. Natl. Acad. Sci. USA, vol. 86(14), 5276-5280) and the blunt ends were ligated into SmαI site of pGEM-7z (Promega). The recombinant plasmid was amplified in E. coli XL1-blue cells and purified. The plasmid was used for sequencing and to prepare a labelled probe for screening of a cDNA library.

Screening of rabbit liver cDNA library in λgt10. The recombinant plasmid containing pGEM-7z and 0.5 kb PCR product (see above) was cut with BamH1 and used to generate a riboprobe (0.5 kb) with the Promega Riboprobe Gemini II Core System. The reaction contained in a total volume of 25 μl:32 mM Tris-HCl, pH 7.5; 5 mM $MgCl_2$; 2 mM spermidine; 8 mM sodium chloride; 8 mM DTT; 40 units RNasin; 0.4 mM of each of ATP, GTP and UTP; 5 μl[α-^{32}P]CTP (800 Ci/mmole); 1 μg of BamH1-cut pGEM-7z/PCR-product recombinant plasmid; and 2 units T7 RNA polymerase. Incubation was at 40°C for 2 hr. RNase-free DNase I (10 units) was added followed by incubation at room temperature for 15 min. Buffer (80 μl of 50 mM Tris-HCl, pH 7.4; 4 mM EDTA; 300 mM NaCl; 0.1% SDS) and tRNA (20 μg) were added followed by extraction with phenol-chloroform-isoamyl alcohol (25:24:1, v/v). The labelled RNA probe was desalted over a Sephadex G-50 column (Nick Column, Pharmacia).

A rabbit liver cDNA library in λgt 10 (5'-stretch, Cat. No TL 1006a from Clontech, EcoRI cloning site) was propagated in E. coli LE392 host cells and 10^6 plaques were screened by standard plaque hybridization techniques (Maniatis et al (1982) Molecular Cloning: a laboratory manual, Cold Spring Harbor, N.Y.:Cold Spring Harbor Laboratory) using the above riboprobe. Following fixation of DNA to nitrocellulose membranes, the membranes were washed for 1 hr at 45°C in 50 mM Tris-HCl, pH 8.0/1 M NaCl/1 mM EDTA/0.1% SDS. Membranes were prehybridized at 50°C for 2 hr in 1M NaCl/50 mM sodium phosphate, pH 6.5/0.1% SDS/50% freshly-deionized formamide/1% glycine/0.5% Blotto/5 mM

EDTA/1% yeast total RNA. Riboprobe (5 x 10^6 cpm/ml hybridization solution) was added and hybridization was carried out at 50°C overnight. Membranes were washed in 2XSSC/0.1% SDS twice for 5 min at room temperature and twice for 15 min at 50°C. Positive isolates were identified by autoradiography and were plaque-purified. DNA was purified from phage lysates, digested with EcoRI, and cDNA inserts were analyzed by agarose gel electrophoresis. The largest cDNA insert obtained was 1.6 kb; it was subcloned into the EcoRI site of pGEM-7z (Promega) by standard methods (Maniatis et al (1982) Molecular Cloning: a laboratory manual, Cold Spring Harbor, N.Y.:Cold Spring Harbor Laboratory) and the recombinant plasmids were transfected into E. coli XL1-blue. Colonies containing the recombinant plasmid were selected and amplified, and plasmid DNA was purified by CsCl gradient centrifugation (Ausubel et al (1990) Current Protocols in Molecular Biology, Media, PA:Greene Publishing Associates and John Wiley and Sons).

The cDNA library was re-screened as described above using a 80 bp riboprobe prepared from the 5'-end of the 1.6 kb clone. The largest cDNA insert obtained was 3.0 kb. This insert was sub-cloned into pGEM-7z as described above and plasmid DNA was purified by CsCl gradient centrifugation (Ausubel et al (1990) Current Protocols in Molecular Biology. Media, PA:Greene Publishing Associates and John Wiley and Sons), to obtain pGEM-7z-rcgnt1.

DNA Sequencing. Two colonies of the pGEM-7z/PCR-product recombinant plasmid (see above) containing inserts in opposite directions were sequenced directly by the single-strand dideoxynucleotide-chain-termination method (Sanger et al, Proc. Natl. Acad. Sci. USA, vol. 74, 5463-5467) using deoxyadenosine 5'-[α-[^{35}S]thio] triphosphate, Sequenase (United States Biochemical) and the forward primer for pGEM-7z. The 1.6 and 3.0 kb clones were sequenced by the Erase-a-Base System (Promega) and the single-strand dideoxynucleotide-chain-termination method. Both DNA strands were sequenced by using colonies in which

the inserts were present in opposite directions. Plasmid DNA (12 μg) was cut with SphI to generate a 5'-overhang and XbaI to generate a 3'-overhang. The cut DNA was digested with exonuclease III (Erase-a-Base System, Promega) for varying lengths of time followed by S1 nuclease digestion. The DNA ends were blunt-ended with the Klenow fragment of E. coli DNA polymerase I and the DNA was circularized with T4 DNA ligase. The ligation mixtures were transfected into competent XL1-blue cells. Miniplasmid preparations were carried out on about 5-10 subclones from each exonuclease III time point and were cut with BamHI and AatII to determine DNA size. Colonies with appropriate deletions were amplified and incubated with M13KO7 helper phage at 37°C for 1 hr followed by amplification in the presence of kanamycin (70 μg/ml) for 6 hr at 37°C. Single-stranded DNA was produced by the helper phage and excreted into the medium. The ss-DNA was purified from the medium by polyethylene glycol precipitation and sequenced by the dideoxynucleotide chain-termination method using deoxyadenosine 5'-[α-[^{35}S]thio]triphosphate, Sequenase (United States Biochemical) and the forward primer for pGEM-7z.

RNA Hybridization. Rabbit liver poly(A)+RNA (5 μg) was denatured in 50% (v/v) formamide/6% (v/v) formaldehyde buffer at 65°C and was resolved by gel electrophoresis in a 1% agarose gel containing 6% (v/v) formaldehyde. The RNA was transferred to a nitrocellulose filter and the filters were hybridized with the ^{32}P-labelled 0.5 kb PCR riboprobe (see above) followed by autoradiography. The specific activity of the probe was about 10^6 dpm/ng and the hybridization solution contained about 10^6 dpm/ml.

In vitro transcription and translation. The recombinant plasmid containing pGEM-7z (Promega) and the 2.5 kb GnT I cDNA insert (rc2500, Figure 2) (pGEM-7z-rcgnt1) was cut with Sph I to generate linear plasmid. RNA was transcribed using the SP6 RNA polymerase promoter and initiation site present in pGEM-7z. RNA synthesis was

carried out at 40°C for 1 hr in a total volume of 50 μl containing 40 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 40 units RNasin (Promega), 0.5 mM of each of ATP, UTP and CTP, 0.1 mM GTP, 0.5 mM $m^7G(5')PPP(5')G$ (Pharmacia), 10 units SP6 RNA polymerase and 10 μg linearized plasmid. Control incubations were carried out in the absence of plasmid or with a linearized pGEM-7z recombinant plasmid containing a non-coding insert. The reaction mixture was extracted twice with phenol-chloroform-isoamyl alcohol (25:24:1, v/v) followed by precipitation with cold ethanol.

Protein synthesis (translation) was carried out at 30°C for 1 hr in a total volume of 50 μl containing all 20 amino acids (1 mM each), 20 units of RNasin, RNA as prepared above, and buffer and rabbit reticulocyte lysate as supplied by Promega (Olliver et al (1984) "In vitro translation of messenger RNA in a rabbit reticulocyte lysate cell-free system", in: M. Walker J., ed., Methods in Molecular Biology, Nucleic Acids, Clifton, N.J.:Humana Press, 145-155). Non-radioactive amino acids were used when the products of translation were assayed for GnT I activity (see below). Separate incubations were carried out with L-[^{35}S]-methionine (1000 Ci/mmole; 90 μCi/incubation) replacing non-radioactive Met; these incubations were analyzed by SDS-polyacrylamide gel electrophoresis followed by autoradiography.

GnT I was assayed (Schachter (1989) Methods Enzymol., vol. 179, 351-396; Brockhausen et al (1988) Biochem. Cell Biol., vol. 66, 1134-1151) in a total volume of 40 μl containing 20 mM $MnCl_2$, bovine serum albumin (1 mg/ml), 0.1% (v/v) Triton X-100, 0.1 M MES (pH 6.1), 0.5 mM UDP-N-[1-^{14}C]acetyl-D-glucosamine (2.2 mCi/mmole), 0.125 M GlcNAc and 0.6 mM Manα1-6(Manα1-3)Manβ-hexyl (a kind gift from Dr. Hans Paulsen, University of Hamburg, Hamburg, Federal Republic of Germany). Incubations were at 37°C for 2 and 16 hr. The reaction was stopped with 0.5 ml 20 mM sodium tetraborate/2 mM EDTA and was passed through a small column of AG1X8

(Cl-form, 100-200 mesh, equilibrated with water) to remove radioactive nucleotide-sugar. The eluate was applied to a Sep-Pak C-18 reverse phase cartridge (Waters) conditioned with 20 ml methanol and 20 ml water. The cartridge was washed with 20 ml water and radioactive product was eluted with 5.0 ml methanol (Palcic et al (1988) Glycoconjuguate J., vol. 5, 49-63). An aliquot was counted directly and the remainder was analyzed by HPLC on a C-18 reverse phase column using acetonitrile-water (12:88) as the mobile phase (Schachter et al (1989) Methods Enzymol., vol. 179, 351-396; Brockhausen et al (1988) Biochem. Cell Biol., vol. 66, 1134-1151). Product co-eluted with a standard preparation of Manα1-6(GlcNAcβ1-2Manα1-3)Manβ-hexyl at 36 min.

Preparation of pGEX-2t-rcgnt1. This plasmid was prepared from pGEM-7z-rcgnt1 by cutting out the insert rcgnt1 with Eco RI. Plasmid pGEX-2t (Pharmacia) was linearized with Eco RI and the insert was ligated into the plasmid by standard procedures. The recombinant plasmid was amplified in E. coli in the presence of ampicillin and purified by cesium chloride centrifugation.

Amplification of cDNA. Three amino acid sequences (Figure 1) were chosen for the design of sense and anti-sense oligonucleotide primers to be used in the PCR amplification of rabbit liver cDNA. Deoxyinosine was substituted in positions where codon degeneracy was >2 (Moremen (1989) Proc. Natl. Acad. Sci. USA, vol. 85(14), 5276-5280); mixed pairs of bases were used in four positions in all three sequences giving a 16-fold mixture of sequences for every primer. Since we had no knowledge of the order of the peptides in the amino acid sequence, PCR was carried out with all six possible combinations of sense and anti-sense primers (2S-3A, 2S-6A, 3S-2A, 3S-6A, 6S-2A, 6S-3A, Figure 1). The products of the PCR reactions were analyzed by agarose gel electrophoresis (Figure 3). Primer-dependent products were obtained with two of the six incubations, i.e., 2S-6A (500 bp) and 3S-6A (450 bp). The complete nucleotide sequence for GnT I is shown in Figure 4.

Oligonucleotide primers 2S and 3A are separated by only nine bases thereby explaining the absence of PCR product with this combination.

Sequence Analysis. The 1.6 kb clone contains 0.5 kb from the 3'-end of the coding region and the full 1.1 kb 3'-untranslated region (rc1600, Figure 2). The 3.0 kb clone yielded a 2485 bp sequence (rc2500, Figure 2; Figure 4). We have shown that subcloning of the 3.0 kb DNA fragment in pGEM-7z results in deletion of a 0.5 kb DNA fragment near the 5'-end of the clone. Comparison of the cDNA sequence shown in Figure 4 with the sequence of human genomic DNA for GnT I (in preparation) has shown that this deleted 0.5 kb DNA fragment is not part of the GnT I gene; we do not know the origin of this DNA.

The GnT I coding sequence has 1341 bp and codes for a membrane-bound protein of 447 amino acids (M_r52,000). There is a single hydrophobic domain (bases 62 to 136) flanked by charged amino acids (Figure 4). Chou-Fasman rules (Chou et al (1978) Adv. Enzymol., vol. 47, 45-147) predict that this hydrophobic segment is capable of propagating an α-helix, as expected for a transmembrane domain.

The presumptive initiation Met codon is at the ATG codon at position 50 which has an A at position 47 thereby fulfilling the requirements for an initiation codon (Kozak (1983) Microbiological Reviews, vol. 47, 1-45). All eight peptides shown in Figure 1 (a total of 103 amino acid residues) can be identified in the sequence (Figure 4); an additional five tentative assignments also match the sequence. GnT I purified from rabbit liver has a molecular weight of about 45 kDa (Nishikawa et al (1988) J. Biol. Chem., vol. 263, 8270-8281). The protein has no N-glycans since none of the nine Asn residues are in a typical Asn-X-Ser(Thr) sequence; we have previously shown that rabbit liver GnT I binds poorly to lectin/agarose columns (Nishikawa et al (1988) J. Biol. Chem., vol. 263, 8270-8281). If there are no or few O-glycans, a

catalytically active protein of 45 kDa can be derived by cleavage at about base position 215 (Figure 4).

Comparison of the GnT I sequence with those of several previously cloned glycosyltransferases (Appert et al (1986) Biochem. Biophys. Res. Commun., vol. 139, 163-168; D'Agostaro et al (1989) Eur. J. Biochem., vol. 183, 211-217; Hollis et al (1989) Biochem. Biophys. Res. Commun., vol. 162, 1069-1075; Joziasse et al (1989) J. Biol. Chem., vol. 264, 14290-14297; Larsen et al (1989) Proc. Natl. Acad. Sci. USA, vol. 86, 8227-8231; Larsen et al (1990) J. Biol. Chem., vol. 265, 7055-7061; Masibay et al (1989) Proc. Natl. Acad. Sci. USA, vol. 86, 5733-5737; Masri et al (1988) Biochem. Biophys. Res. Commun., vol. 157, 657-663; Narimatsu et al (1986) Proc. Nat. Acad. Sci. USA, vol. 83, 4720-4724; Russo et al (1990) J. Biol. Chem., vol. 265, 3324-3331; Shaper et al (1986) Proc. Nat. Acad. Sci. USA, vol. 83, 1573-1577; Shaper et al (1988) J. Biol. Chem., vol. 263, 10420-10428; Shaper et al (1988) Biochemie., vol. 70, 1683-1688; Shaper et al (1990) Proc. Natl. Acad. Sci. USA, vol. 87, 791-795; Smith et al (1990) J. Biol. Chem., vol. 265, 6225-6234; Weinstein et al (1987) J. Biol. Chem., vol. 262, 17735-17743) revealed no sequence homology but GnT I appears to have a domain structure typical of these enzymes (Paulson et al (1989) J. Biol. Chem., vol. 264, 17615-17618). Searches of the GenBank nucleotide data base (release 62.0) with the coding region of GnT I and of the PIR Protein Data Base (release 23.0) with the GnT I amino acid sequence revealed no significant similarities to other sequences.

The complete sequence has a long 3'-untranslated region (bases 1391-2479) containing the consensus polyadenylation signal AATAAA at position 2435 (Tosi et al (1981) Nucleic Acids Research, vol. 9, 2313-2323). Long 3'-untranslated regions are typical of the known glycosyltransferase genes and may be a feature present in other Golgi-localized enzymes (Moremen (1989) Proc. Natl. Acad. Sci. USA, vol. 86(14), 5276-5280).

Northern Blot Analysis. The PCR riboprobe was used to determine the size of mRNA in rabbit liver. A major band was detected at about 3.0 kb with some smearing at lower molecular weights (data not shown) indicating that the 2.5 kb cDNA clone (Figure 4) may not be full-length.

In Vitro transcription and translation. Transcription of the linearized pGEM-7z/2.5 kb GnT I cDNA recombinant plasmid (pGEM-7z-rcgnt1) followed by translation in the presence of L-[^{35}S]Met resulted in the appearance of a strong radioactive 52 kDa band on SDS-polyacrylamide gel electrophoresis; this band was not seen in control incubations lacking plasmid or containing control plasmid (Figure 5). The molecular weight matches the prediction for the open reading frame shown in Figure 4. Table 1 shows the results of GnT I assays carried out on the transcription-translation incubations. The incubation containing the pGEM-7z/2.5 kb GnT I cDNA recombinant plasmid (pGEM-7z-rcgnt1) has appreciable GnT I activity whereas both controls show low activity. It is concluded that the 2.5 kb sequence shown in Figure 4 can code for the synthesis of catalytically active GnT I.

TABLE 1

In vitro transcription-translation of rabbit GnT I cDNA

Conditions of transcription	GnT I product (nmoles/total transcription incubation)		
	Sep-Pak assays 2 hr	Sep-Pak assays 16 hr	HPLC assays 16 hr
No plasmid	0.04	0.21	--
Control Plasmid	0.04	0.21	0.29
2.5 kb GnT I cDNA (pGEM-7z-rcgnt1)	0.41	1.05	1.32

II. Human GnT I:

The polymerase chain reaction (PCR) was used to obtain a 0.5 kb ds-cDNA representing the carboxy terminal half of the rabbit liver GnT I coding sequence and labelled this DNA fragment by the random primer technique. The preparation of this probe is described above.

The rabbit cDNA probe was used to screen 10^6 plaques from an amplified human genomic DNA library in λEMBL3 prepared from chromosomal DNA from chronic myeloid leukemia cells. Positive plaques (23) were purified and phage DNA was subjected to restriction enzyme analysis using the 0.5 kb rabbit cDNA as probe. All 23 preparations gave the same Sau3A 0.4 kb fragment. This fragment showed 87% base similarity and 90% amino acid sequence similarity to the rabbit GnT I carboxy-terminal sequence. Inserts of 13 and 15 kb were cut from two of the human genomic DNA clones with SAII and subcloned into plasmid pGEM-5zf(+)(Promega). Restriction maps of the two inserts show that they represent an over-lapping 18 kb DNA sequence.

The coding sequence was located in a 4.0 kb fragment of human genomic DNA by screening restriction maps with a probe containing the entire coding region of the rabbit GnT I cDNA. This 4.0 kb DNA fragment was cut out by restriction enzymes and subcloned into the sequencing vector pGEM-5zf(+) to yield pGEM-5z-hggnt1 and sequenced. Transfection of the gene into Lec 1 Chinese hamster ovary cell mutants (which lack GnT I activity) results in the expression of GnT I activity indicating the presence of a functional promoter 5'-upstream of the transcription start site.

The 4 kb sequence contains an open reading frame coding for a protein with 445 amino acids (2 less than the rabbit enzyme). The DNA contains a functional promoter and an intronless gene. The similarity between the rabbit and human enzymes is 85% for the nucleotide coding sequences and over 90% for the amino acid sequences.

Obviously, numerous modifications and variations of the present invention are possible in light of the above

teachings. It is therefore to be understood that, within the scope of the appended claims, the invention may be practiced otherwise than as specifically described herein. The references cited in the specification are incorporated herein by reference.

THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. An isolated DNA sequence encoding a protein having the amino acid sequence of formula I:

MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE
LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR
ARG PRO VAL PRO SER ARG LEU PRO SER ASP ASN ALA LEU ASP ASP
ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP
ALA GLU VAL GLU LEU GLU ARG GLN ARG GLY LEU LEU GLN GLN ILE
ARG GLU HIS HIS ALA LEU TRP SER GLN ARG TRP LYS VAL PRO THR
ALA ALA PRO PRO ALA GLN PRO HIS VAL PRO VAL THR PRO PRO PRO
ALA VAL ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL
ARG ARG CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU
LEU PHE PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR
ALA GLN VAL ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG
GLN PRO ASP LEU SER ASN ILE ALA VAL GLN PRO ASP HIS ARG LYS
PHE GLN GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU
GLY GLN ILE PHE HIS ASN PHE ASN TYR PRO ALA ALA VAL VAL VAL
GLU ASP ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE GLN
ALA THR TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL
SER ALA TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP SER SER
LYS PRO GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY
TRP LEU LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP
PRO LYS ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG
LYS GLY ARG ALA CYS VAL ARG PRO GLU ILE SER ARG THR MET THR
PHE GLY ARG LYS GLY VAL SER HIS GLY GLN PHE PHE ASP GLN HIS
LEU LYS PHE ILE LYS LEU ASN GLN GLN PHE VAL PRO PHE THR GLN
LEU ASP LEU SER TYR LEU GLN GLN GLU ALA TYR ASP ARG ASP PHE
LEU ALA ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL
ARG THR ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR
THR GLY ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL
MET ASP ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY
ILE VAL THR PHE LEU PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO
PRO GLN THR TRP ASP GLY TYR ASP PRO SER TRP THR.

2. The DNA sequence of Claim 1, having the nucleotide sequence of formula III:

atg ctg aag aag cag tct gct ggg ctt gtg ctg tgg ggt gct atc
ctc ttt gtg gcc tgg aat gcc ctg ctg ctc ctc ttc ttc tgg aca

cgt cca gtg cct agc agg ctg ccg tca gac aat gct ctc gat gat
gac cct gcc agc ctc acc cgt gag gtg atc cgc tta gct cag gat
gcc gag gta gag ttg gaa cgt cag cgg gga ctg ttg cag cag att
agg gag cac cat gct ctt tgg agc cag cgg tgg aag gtg cct act
gca gcc cct cct gct cag ccg cat gtg cct gtg acc cca ccg cca
gct gtg atc ccc atc ctg gta att gcc tgt gac cgc agc acc gtc
cgc cgc tgt ttg gac aag cta ctg cat tat cgg cct tca gct gag
ctg ttc ccc atc att gtc agc cag gac tgt ggg cat gag gag aca
gcc cag gtc att gct tcc tat ggc agc gca gtc aca cac atc cgg
caa cct gac ctg agc aac att gct gtg cag ccc gac cac cgc aag
ttc cag ggc tac tac aag atc gca cgg cat tac cgc tgg gca ttg
ggc caa atc ttc cac aat ttc aac tac cca gca gct gtg gtg gtg
gag gat gat ctc gag gtg gca cca gac ttc ttt gag tac ttc cag
gcc act tac cca ctg ttg aaa gca gac ccc tcc ctc tgg tgt gtg
tct gcc tgg aat gac aat ggc aaa gaa cag atg gta gac tcg agt
aag cca gag tta ctc tac cgc aca gat ttc ttt cct ggc tta ggc
tgg tta ctg ttg gct gaa ctc tgg gct gaa ctg gag ccc aag tgg
ccc aaa gcc ttc tgg gat gac tgg atg cgc cgg cct gag cag cga
aag ggg agg gcc tgt gtg cgt cca gaa atc tca aga aca atg aca
ttt ggc cgg aag ggt gtg agc cat ggg cag ttc ttt gac cag cat
ctc aag ttc atc aag ctg aac cag cag ttt gta ccc ttc acc cag
ctg gac ctg tcg tac ctt cag cag gag gcc tat gac cgg gat ttc
ctt gct cgt gtt tat ggt gct ccc cag tta cag gtg gag aaa gtg
agg acc aat gac cgg aag gag cta gga gag gtg cgc gta cag tac
aca ggc agg gac agc ttc aag gct ttc gcc aag gcc ctg ggt gtc
atg gat gac ctc aaa tca ggt gta ccc agg gct gga tac cgg ggc
att gtc acc ttc tta ttc cgg ggc cgc cgt gtc cac ctg gcg ccc
cct cag act tgg gat ggc tat gat cct agt tgg act.

3. The DNA sequence of Claim 1, having the nucleotide sequence of formula IV:

gaattccggc aagtcatacc tttgcctgcc ctcccctgtg ggggccagg
atg ctg aag aag cag tct gct ggg ctt gtg ctg tgg ggt gct atc
ctc ttt gtg gcc tgg aat gcc ctg ctg ctc ctc ttc ttc tgg aca
cgt cca gtg cct agc agg ctg ccg tca gac aat gct ctc gat gat
gac cct gcc agc ctc acc cgt gag gtg atc cgc tta gct cag gat
gcc gag gta gag ttg gaa cgt cag cgg gga ctg ttg cag cag att
agg gag cac cat gct ctt tgg agc cag cgg tgg aag gtg cct act

gca gcc cct cct gct cag ccg cat gtg cct gtg acc cca ccg cca
gct gtg atc ccc atc ctg gta att gcc tgt gac cgc agc acc gtc
cgc cgc tgt ttg gac aag cta ctg cat tat cgg cct tca gct gag
ctg ttc ccc atc att gtc agc cag gac tgt ggg cat gag gag aca
gcc cag gtc att gct tcc tat ggc agc gca gtc aca cac atc cgg
caa cct gac ctg agc aac att gct gtg cag ccc gac cac cgc aag
ttc cag ggc tac tac aag atc gca cgg cat tac cgc tgg gca ttg
ggc caa atc ttc cac aat ttc aac tac cca gca gct gtg gtg gtg
gaa gat gat ctc gag gtg gca cca gac ttc ttt gag tac ttc cag
gcc act tac cca ctg ttg aaa gca gac ccc tcc ctc tgg tgt gtg
tct gcc tgg aat gac aat ggc aaa gaa cag atg gta gac tcg agt
aag cca gag tta ctc tac cgc aca gat ttc ttt cct ggc tta ggc
tgg tta ctg ttg gct gaa ctc tgg gct gaa ctg gag ccc aag tgg
ccc aaa gcc ttc tgg gat gac tgg atg cgc cgg cct gag cag cga
aag ggg agg gcc tgt gtg cgt cca gaa atc tca aga aca atg aca
ttt ggc cgg aag ggt gtg agc cat ggg cag ttc ttt gac cag cat
ctc aag ttc atc aag ctg aac cag cag ttt gta ccc ttc acc cag
ctg gac ctg tcg tac ctt cag cag gag gcc tat gac cgg gat ttc
ctt gct cgt gtt tat ggt gct ccc cag tta cag gtg gag aaa gtg
agg acc aat gac cgg aag gag cta gga gag gtg cgc gta cag tac
aca ggc agg gac agc ttc aag gct ttc gcc aag gcc ctg ggt gtc
atg gat gac ctc aaa tca ggt gta ccc agg gct gga tac cgg ggc
att gtc acc ttc tta ttc cgg ggc cgc cgt gtc cac ctg gcg ccc
cct cag act tgg gat ggc tat gat cct agt tgg act

taacagctcc tgcctgtccc ttctgggctc cttccttgca atttcatgat ctaagatggg
accgtagtcc ctgggctgca ttgtcttttc tgtctttccc tcttgggtcc attttttttt
ttttcttttt tgagtggcat ttgaatacac agatgacaag gtgagggttc ttttgttaaa
ggagttagat cagggaaagc attctgctgt ctgttgggta tcaagcagca aaccactgtg
tgatagggga agaatgggct ttttggggcc agaaatatcc atgttctgag ttttttcttt
aggtcatctg cagaggagtt ggcaacttta gctttcttaa ccaggccttt tctttctgac
ctgagagcca gggcatgaga cttcttgttc atgctccttt ttaccttccc ctaataaggg
tctgggctac aggagaagtg aacatattgt ggccagaata atactaacca gaggggcctc
attgtcagag tctaggtgca gttattgggt tgtcagagtt aatgccttct gttcttcttt
ccttattcct gacttctgtc agctcttctt tctttgcagc ctagcaattt ttggttctaa
gatgaaaaat gaagaggaaa agaaatattc gcacccagct attgggagaa aggtagtggg
aaaaaaactt cattgtacca cttcaaagag acactcttga cctcttcctt tctaaaaatt
agtcccctcc ctgttgcttc aggagaatgc tgtgctggtc agttctgtgt gatccttctt
ccctgagttt tatacacagg ctcctcccta aggctgtggc ttctggtggc cctcctgaca
taagttacag tggccaagac caggacaact ccggccatga gctaagtcct gcctaccttc
tccaaaacat tcccatgtcc tcacaggcta ggatgcagat gttggttgga gaggaatttg
tgtgtgtgtg tgtgtgtgtg tgtgttttct tgcctgacct cagtttcatg gatgaaaagt
ggaagctaca gaattatttt caaaaataaa ggctgaattg tctgaaaaaa aaaaaaaaaa
aaaaaaccgg aattc.

4. An isolated DNA sequence encoding a protein having the amino acid sequence of formula II:

MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE
LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR
ARG PRO ALA PRO GLY ARG PRO PRO SER VAL SER ALA LEU ASP GLY
ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP
ALA GLU VAL GLU LEU GLU ARG ARG ARG GLY LEU LEU GLN GLN ILE
GLY ASP ALA LEU SER SER GLN ARG GLY ARG VAL PRO THR ALA ALA
PRO PRO ALA GLN PRO ARG VAL PRO VAL THR PRO ALA PRO ALA VAL
ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL ARG ARG
CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU LEU PHE
PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR ALA GLN
ALA ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG GLN PRO
ASP LEU SER SER ILE ALA VAL PRO PRO ASP HIS ARG LYS PHE GLN
GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU GLY GLN
VAL PHE ARG GLN PHE ARG PHE PRO ALA ALA VAL VAL VAL GLU ASP
ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE ARG ALA THR
TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL SER ALA
TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP ALA SER ARG PRO
GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY TRP LEU
LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP PRO LYS
ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG GLN GLY
ARG ALA CYS ILE ARG PRO GLU ILE SER ARG THR MET THR PHE GLY
ARG LYS GLY VAL THR HIS GLY GLN PHE PHE ASP GLN HIS LEU LYS
PHE ILE LYS LEU ASN GLN GLN PHE VAL HIS PHE THR GLN LEU ASP
LEU SER TYR LEU GLN ARG GLU ALA TYR ASP ARG ASP PHE LEU ALA
ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL ARG THR
ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR THR GLY
ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL MET ASP
ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY ILE VAL
THR PHE GLN PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO PRO PRO.

5. The DNA sequence of Claim 4, having the nucleotide sequence of formula V:

atgctgaa gaagcagtct gcagggcttg tgctgtgggg cgctatcctc tttgtggcct
ggaatgccct gctgctcctc ttcttctgga cgcgcccagc acctggcagg ccaccctcag
tcagcgctct cgatggcgac cccgccagcc tcacccggga agtgattcgc ctggcccaag
acgccgaggt ggagctggag cgcaggcgtg ggctgctgca gcagatcggg gatgccctgt
cgagccagcg ggggagggtg cccaccgcgg cccctcccgc ccagccgcgt gtgcctgtga
cccccgcgcc ggcggtgatt cccatcctgg tcatcgcctg tgaccgcagc actgttcggc

gctgcctgga caagctgctg cattatcggc cctcggctga gctcttcccc atcatcgtta
gccaggactg cgggcacgag gagacggccc aggccatcgc ctcctacggc agcgcggtca
cgcacatccg gcagcccgac ctgagcagca ttgcggtgcc gccggaccac cgcaagttcc
agggctacta caagatcgcg cgccactacc gctgggcgct gggccaggtc ttccggcagt
ttcgcttccc cgcggccgtg gtggtggagg atgacctgga ggtggccccg gacttcttcg
agtactttcg ggccacctat ccgctgctga aggccgaccc ctccctgtgg tgcgtctcgg
cctggaatga caacggcaag gagcagatgg tggacgccag caggcctgag ctgctctacc
gcaccgactt tttccctggc ctgggctggc tgctgttggc cgagctctgg gctgagctgg
agcccaagtg gccaaaggcc ttctgggacg actggatgcg gcggccggag cagcggcagg
ggcgggcctg catacgccct gagatctcaa gaacgatgac ctttggccgc aagggtgtga
cgcacgggca gttctttgac cagcacctca agtttatcaa gctgaaccag cagtttgtgc
acttcaccca gctggacctg tcttacctgc agcgggaggc ctatgaccga gatttcctcg
cccgcgtcta cggtgctccc cagctgcagg tggagaaagt gaggaccaat gaccggaagg
agctggggga ggtgcgggtg cagtatacgg ggagggacag cttcaaggct ttcgccaagg
ctctgggtgt tatggatgac cttaagtcgg gggttccgag agctggctac cggggtattg
tcaccttcca gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct
atgatcctag ctggaat.

6. The DNA sequence of Claim 4, having the nucleotide sequence of formula VI:

aagttttgaa tgtttaagtt tatttaagtt tatttctaaa tatttttcta tttctctggc
ttttgtaagt agggttttct catccatgtt ttcttctcat gagttatttg tggatatgaa
ggctatccat tagtatatgt tgatttttat attacacttc cttgctcagt tcattattga
ttctttttga gttttccagg catattctca caagtaaaga taatagaaat agtttgcttc
ctttccactt ctgctttgaa tttttttttc ttggttcatt tgcattggct gcttcctcca
gcaaaatgtt aaataaccct ggagatgatg ggcaacttcg ttttgctcct gacattcgtg
gggtgcctct ggtgcttccc tgttggtaag gggttaactg tagccctgag gtgggacatt
tgattttaaa aatcagtcat cttggggcgc ttaggttaga ggaatggtag gcagatgctg
tcactccttg cccctcccct cctccttccc acctggaggg gaaatgaaat ctgacaggta
gaaagagggg agttggggtt ctttttctct ctcccctccac cagcatcact ctctgcctct
ccctcaaaaa tacgttcctg ggtcaggata tatgttgact ccctagagag ctctggagtc
aacctcctgg ccttcctcca ccctcactct tggccttttc ctgcccccat ttcctctacc
tgtggggcat ggagccacga gcctttgtgt gacggtttgc tttctctctc ctgtctttag
gtgcatggct gcctcctaat cccatagtcc agaggaggca tccctaggac tgcgggcaag
ggagccgcaa gcccagggca gccttgaacc gtcccctggc ctgccctccg gtgggggcca
ggatgctgaa gaagcagtct gcagggcttg tgctgtgggg cgctatcctc tttgtggcct
ggaatgccct gctgctcctc ttcttctgga cgcgcccagc acctggcagg ccaccctcag
tcagcgctct cgatggcgac cccgccagcc tcacccggga agtgattcgc ctggcccaag
acgccgaggt ggagctggag cgcaggcgtg ggctgctgca gcagatcggg gatgccctgt
cgagccagcg ggggagggtg cccaccgcgg cccctcccgc ccagccgcgt gtgcctgtga
cccccgcgcc ggcggtgatt cccatcctgg tcatcgcctg tgaccgcagc actgttcggc
gctgcctgga caagctgctg cattatcggc cctcggctga gctcttcccc atcatcgtta
gccaggactg cgggcacgag gagacggccc aggccatcgc ctcctacggc agcgcggtca
cgcacatccg gcagcccgac ctgagcagca ttgcggtgcc gccggaccac cgcaagttcc
agggctacta caagatcgcg cgccactacc gctgggcgct gggccaggtc ttccggcagt
ttcgcttccc cgcggccgtg gtggtggagg atgacctgga ggtggccccg gacttcttcg
agtactttcg ggccacctat ccgctgctga aggccgaccc ctccctgtgg tgcgtctcgg
cctggaatga caacggcaag gagcagatgg tggacgccag caggcctgag ctgctctacc
gcaccgactt tttccctggc ctgggctggc tgctgttggc cgagctctgg gctgagctgg
agcccaagtg gccaaaggcc ttctgggacg actggatgcg gcggccggag cagcggcagg
ggcgggcctg catacgccct gagatctcaa gaacgatgac ctttggccgc aagggtgtga
cgcacgggca gttctttgac cagcacctca agtttatcaa gctgaaccag cagtttgtgc
acttcaccca gctggacctg tcttacctgc agcgggaggc ctatgaccga gatttcctcg
cccgcgtcta cggtgctccc cagctgcagg tggagaaagt gaggaccaat gaccggaagg
agctggggga ggtgcgggtg cagtatacgg ggagggacag cttcaaggct ttcgccaagg

ctctgggtgt tatggatgac cttaagtcgg gggttccgag agctggctac cggggtattg
tcaccttcca gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct
atgatcctag ctggaattag cacctgcctg tccttcctgg gcccccttctt gccacatcat
gagctgaggt gaccacagtc cccaggctgc atcggcctgc ctgtgtttcc ctcttaggtg
catttatctt tttgattttt ccgagtggca tttaagtgca caaatgataa caagaggatt
attctcccgt tctcaaggga gtcagatcag gggaactatt ctagggtatg ttgcggggta
ttaagcagga aaacactgtg tggtgggggg cactgggctt gttggggcca caaatgtcca
cgtcctgagc tttctcctgg agcatgtgca gagagtttgg caacgttcgc tctcttgacc
agacccccttc tccctgactg gctcttccag ccaggcacga gccctccttc tatacctgct
ccccttccca gtggggactg agttatggga gaaggggaca tatttgtggc caaaatgata
ctaaccaaag gggcttcctt gtcagggcct ggtggagttg gtgggtcatc ggggctcact
gcctcctgcc cttctctcct gtctgacccc cacttagccc ttctctcctt gcagcctagc
agtttatagt tctgagatgg aaagttgaag ggggcaagca agacctctcc tcagcccatg
cccagctgtc aggagagagg tgcagggagg aaggccttgt gctgggacaa cctctctctt
gccttacctt cagagaggac tatgccctga cccctccttt ctgaaaatca gtgccctccc
tgttgctcta ggaggctcct gctggcttgg tagaagacag aattcgatct gcctgtccct
ttttcccctg gggtttgaca cacaggctcc tctcagcatg aggtggagca gtgaccaggt
ggagcagtga ccaggacgcc tctggcccag tgctgcccag cctccccgcc cgctcccagg
cgccccatgt cctcacaggc caggacgcca tggcggccgg gagcatgcga.

7. A plasmid, comprising a DNA sequence encoding a protein having the amino acid sequence of formula I:

MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE
LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR
ARG PRO VAL PRO SER ARG LEU PRO SER ASP ASN ALA LEU ASP ASP
ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP
ALA GLU VAL GLU LEU GLU ARG GLN ARG GLY LEU LEU GLN GLN ILE
ARG GLU HIS HIS ALA LEU TRP SER GLN ARG TRP LYS VAL PRO THR
ALA ALA PRO PRO ALA GLN PRO HIS VAL PRO VAL THR PRO PRO PRO
ALA VAL ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL
ARG ARG CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU
LEU PHE PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR
ALA GLN VAL ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG
GLN PRO ASP LEU SER ASN ILE ALA VAL GLN PRO ASP HIS ARG LYS
PHE GLN GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU
GLY GLN ILE PHE HIS ASN PHE ASN TYR PRO ALA ALA VAL VAL VAL
GLU ASP ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE GLN
ALA THR TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL
SER ALA TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP SER SER
LYS PRO GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY
TRP LEU LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP
PRO LYS ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG
LYS GLY ARG ALA CYS VAL ARG PRO GLU ILE SER ARG THR MET THR
PHE GLY ARG LYS GLY VAL SER HIS GLY GLN PHE PHE ASP GLN HIS

LEU LYS PHE ILE LYS LEU ASN GLN GLN PHE VAL PRO PHE THR GLN LEU ASP LEU SER TYR LEU GLN GLN GLU ALA TYR ASP ARG ASP PHE LEU ALA ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL ARG THR ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR THR GLY ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL MET ASP ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY ILE VAL THR PHE LEU PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO PRO GLN THR TRP ASP GLY TYR ASP PRO SER TRP THR.

8. The plasmid of Claim 7, wherein said DNA sequence has the formula III:

atg ctg aag aag cag tct gct ggg ctt gtg ctg tgg ggt gct atc ctc ttt gtg gcc tgg aat gcc ctg ctg ctc ctc ttc ttc tgg aca cgt cca gtg cct agc agg ctg ccg tca gac aat gct ctc gat gat gac cct gcc agc ctc acc cgt gag gtg atc cgc tta gct cag gat gcc gag gta gag ttg gaa cgt cag cgg gga ctg ttg cag cag att agg gag cac cat gct ctt tgg agc cag cgg tgg aag gtg cct act gca gcc cct cct gct cag ccg cat gtg cct gtg acc cca ccg cca gct gtg atc ccc atc ctg gta att gcc tgt gac cgc agc acc gtc cgc cgc tgt ttg gac aag cta ctg cat tat cgg cct tca gct gag ctg ttc ccc atc att gtc agc cag gac tgt ggg cat gag gag aca gcc cag gtc att gct tcc tat ggc agc gca gtc aca cac atc cgg caa cct gac ctg agc aac att gct gtg cag ccc gac cac cgc aag ttc cag ggc tac tac aag atc gca cgg cat tac cgc tgg gca ttg ggc caa atc ttc cac aat ttc aac tac cca gca gct gtg gtg gtg gag gat gat ctc gag gtg gca cca gac ttc ttt gag tac ttc cag gcc act tac cca ctg ttg aaa gca gac ccc tcc ctc tgg tgt gtg tct gcc tgg aat gac aat ggc aaa gaa cag atg gta gac tcg agt aag cca gag tta ctc tac cgc aca gat ttc ttt cct ggc tta ggc tgg tta ctg ttg gct gaa ctc tgg gct gaa ctg gag ccc aag tgg ccc aaa gcc ttc tgg gat gac tgg atg cgc cgg cct gag cag cga aag ggg agg gcc tgt gtg cgt cca gaa atc tca aga aca atg aca ttt ggc cgg aag ggt gtg agc cat ggg cag ttc ttt gac cag cat ctc aag ttc atc aag ctg aac cag cag ttt gta ccc ttc acc cag ctg gac ctg tcg tac ctt cag cag gag gcc tat gac cgg gat ttc ctt gct cgt gtt tat ggt gct ccc cag tta cag gtg gag aaa gtg agg acc aat gac cgg aag gag cta gga gag gtg cgc gta cag tac aca ggc agg gac agc ttc aag gct ttc gcc aag gcc ctg ggt gtc

atg gat gac ctc aaa tca ggt gta ccc agg gct gga tac cgg ggc
att gtc acc ttc tta ttc cgg ggc cgc cgt gtc cac ctg gcg ccc
cct cag act tgg gat ggc tat gat cct agt tgg act.

9. The plasmid of Claim 7, wherein said DNA sequence has the formula IV:

gaattccggc aagtcatacc tttgcctgcc ctcccctgtg ggggccagg
atg ctg aag aag cag tct gct ggg ctt gtg ctg tgg ggt gct atc
ctc ttt gtg gcc tgg aat gcc ctg ctg ctc ctc ttc ttc tgg aca
cgt cca gtg cct agc agg ctg ccg tca gac aat gct ctc gat gat
gac cct gcc agc ctc acc cgt gag gtg atc cgc tta gct cag gat
gcc gag gta gag ttg gaa cgt cag cgg gga ctg ttg cag cag att
agg gag cac cat gct ctt tgg agc cag cgg tgg aag gtg cct act
gca gcc cct cct gct cag ccg cat gtg cct gtg acc cca ccg cca
gct gtg atc ccc atc ctg gta att gcc tgt gac cgc agc acc gtc
cgc cgc tgt ttg gac aag cta ctg cat tat cgg cct tca gct gag
ctg ttc ccc atc att gtc agc cag gac tgt ggg cat gag gag aca
gcc cag gtc att gct tcc tat ggc agc gca gtc aca cac atc cgg
caa cct gac ctg agc aac att gct gtg cag ccc gac cac cgc aag
ttc cag ggc tac tac aag atc gca cgg cat tac cgc tgg gca ttg
ggc caa atc ttc cac aat ttc aac tac cca gca gct gtg gtg gtg
gaa gat gat ctc gag gtg gca cca gac ttc ttt gag tac ttc cag
gcc act tac cca ctg ttg aaa gca gac ccc tcc ctc tgg tgt gtg
tct gcc tgg aat gac aat ggc aaa gaa cag atg gta gac tcg agt
aag cca gag tta ctc tac cgc aca gat ttc ttt cct ggc tta ggc
tgg tta ctg ttg gct gaa ctc tgg gct gaa ctg gag ccc aag tgg
ccc aaa gcc ttc tgg gat gac tgg atg cgc cgg cct gag cag cga
aag ggg agg gcc tgt gtg cgt cca gaa atc tca aga aca atg aca
ttt ggc cgg aag ggt gtg agc cat ggg cag ttc ttt gac cag cat
ctc aag ttc atc aag ctg aac cag cag ttt gta ccc ttc acc cag
ctg gac ctg tcg tac ctt cag cag gag gcc tat gac cgg gat ttc
ctt gct cgt gtt tat ggt gct ccc cag tta cag gtg gag aaa gtg
agg acc aat gac cgg aag gag cta gga gag gtg cgc gta cag tac
aca ggc agg gac agc ttc aag gct ttc gcc aag gcc ctg ggt gtc
atg gat gac ctc aaa tca ggt gta ccc agg gct gga tac cgg ggc
att gtc acc ttc tta ttc cgg ggc cgc cgt gtc cac ctg gcg ccc
cct cag act tgg gat ggc tat gat cct agt tgg act

taacagctcc tgcctgtccc ttctgggctc cttccttgca atttcatgat ctaagatggg
accgtagtcc ctgggctgca ttgtcttttc tgtctttccc tcttgggtcc attttttttt
ttttcttttt tgagtggcat ttgaatacac agatgacaag gtgagggttc ttttgttaaa
ggagttagat cagggaaagc attctgctgt ctgttgggta tcaagcagca aaccactgtg
tgatagggga agaatgggct ttttggggcc agaaatatcc atgttctgag ttttttcttt
aggtcatctg cagaggagtt ggcaacttta gctttcttaa ccaggccttt tctttctgac
ctgagagcca gggcatgaga cttcttgttc atgctccttt ttaccttccc ctaataaggg
tctgggctac aggagaagtg aacatattgt ggccagaata atactaacca gaggggcctc
attgtcagag tctaggtgca gttattgggt tgtcagagtt aatgccttct gttcttcttt
ccttattcct gacttctgtc agctcttctt tctttgcagc ctagcaattt ttggttctaa
gatgaaaaat gaagaggaaa agaaatattc gcacccagct attgggagaa aggtagtggg
aaaaaaactt cattgtacca cttcaaagag acactcttga cctcttcctt tctaaaaatt
agtcccctcc ctgttgcttc aggagaatgc tgtgctggtc agttctgtgt gatccttctt
ccctgagttt tatacacagg ctcctcccta aggctgtggc ttctggtggc cctcctgaca
taagttacag tggccaagac caggacaact ccggccatga gctaagtcct gcctaccttc
tccaaaacat tcccatgtcc tcacaggcta ggatgcagat gttggttgga gaggaatttg
tgtgtgtgtg tgtgtgtgtg tgtgttttct tgcctgacct cagtttcatg gatgaaaagt
ggaagctaca gaattatttt caaaaataaa ggctgaattg tctgaaaaaa aaaaaaaaaa
aaaaaaccgg aattc.

10. A plasmid, comprising a DNA sequence encoding a protein having the amino acid sequence of formula II:

MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE
LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR
ARG PRO ALA PRO GLY ARG PRO PRO SER VAL SER ALA LEU ASP GLY
ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP
ALA GLU VAL GLU LEU GLU ARG ARG ARG GLY LEU LEU GLN GLN ILE
GLY ASP ALA LEU SER SER GLN ARG GLY ARG VAL PRO THR ALA ALA
PRO PRO ALA GLN PRO ARG VAL PRO VAL THR PRO ALA PRO ALA VAL
ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL ARG ARG
CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU LEU PHE
PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR ALA GLN
ALA ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG GLN PRO
ASP LEU SER SER ILE ALA VAL PRO PRO ASP HIS ARG LYS PHE GLN
GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU GLY GLN
VAL PHE ARG GLN PHE ARG PHE PRO ALA ALA VAL VAL VAL GLU ASP
ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE ARG ALA THR
TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL SER ALA
TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP ALA SER ARG PRO
GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY TRP LEU
LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP PRO LYS
ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG GLN GLY
ARG ALA CYS ILE ARG PRO GLU ILE SER ARG THR MET THR PHE GLY
ARG LYS GLY VAL THR HIS GLY GLN PHE PHE ASP GLN HIS LEU LYS

PHE ILE LYS LEU ASN GLN GLN PHE VAL HIS PHE THR GLN LEU ASP LEU SER TYR LEU GLN ARG GLU ALA TYR ASP ARG ASP PHE LEU ALA ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL ARG THR ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR THR GLY ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL MET ASP ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY ILE VAL THR PHE GLN PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO PRO PRO.

11. The plasmid of Claim 10, wherein said DNA sequence has the formula V:

atgctgaa gaagcagtct gcagggcttg tgctgtgggg cgctatcctc tttgtggcct
ggaatgccct gctgctcctc ttcttctgga cgcgcccagc acctggcagg ccaccctcag
tcagcgctct cgatggcgac cccgccagcc tcacccggga agtgattcgc ctggcccaag
acgccgaggt ggagctggag cgcaggcgtg ggctgctgca gcagatcggg gatgccctgt
cgagccagcg ggggagggtg cccaccgcgg ccccteccgc ccagccgcgt gtgcctgtga
cccccgcgcc ggcggtgatt cccatcctgg tcatcgcctg tgaccgcagc actgttcggc
gctgcctgga caagctgctg cattatcggc cctcggctga gctcttcccc atcatcgtta
gccaggactg cgggcacgag gagacggccc aggccatcgc ctcctacggc agcgcggtca
cgcacatccg gcagcccgac ctgagcagca ttgcggtgcc gccggaccac cgcaagttcc
agggctacta caagatcgcg cgccactacc gctgggcgct gggccaggtc ttccggcagt
ttcgcttccc cgcggccgtg gtggtggagg atgacctgga ggtggccccg gacttcttcg
agtactttcg ggccacctat ccgctgctga aggccgaccc ctccctgtgg tgcgtctcgg
cctggaatga caacggcaag gagcagatgg tggacgccag caggcctgag ctgctctacc
gcaccgactt tttccctggc ctgggctggc tgctgttggc cgagctctgg gctgagctgg
agcccaagtg gccaaaggcc ttctgggacg actggatgcg gcggccggag cagcggcagg
ggcgggcctg catacgccct gagatctcaa gaacgatgac ctttggccgc aagggtgtga
cgcacgggca gttctttgac cagcacctca agtttatcaa gctgaaccag cagtttgtgc
acttcaccca gctggacctg tcttacctgc agcgggaggc ctatgaccga gatttcctcg
cccgcgtcta cggtgctccc cagctgcagg tggagaaagt gaggaccaat gaccggaagg
agctgggggа ggtgcgggtg cagtatacgg ggagggacag cttcaaggct ttcgccaagg
ctctgggtgt tatggatgac cttaagtcgg gggttccgag agctggctac cggggtattg
tcaccttcca gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct
atgatcctag ctggaat.

12. The plasmid of Claim 10, wherein said DNA sequence has the formula VI:

aagttttgaa tgtttaagtt tatttaagtt tatttctaaa tattttctca tttctctggc
ttttgtaagt agggttttct catccatgtt ttcttctcat gagttatttg tggatatgaa
ggctatccat tagtatatgt tgatttttat attacacttc cttgctcagt tcattattga
ttctttttga gttttccagg catattctca caagtaaaga taatagaaat agtttgcttc
ctttccactt ctgctttgaa tttttttttc ttggttcatt tgcattggct gcttcctcca
gcaaaatgtt aaataaccct ggagatgatg ggcaacttcg ttttgctcct gacattcgtg
gggtgcctct ggtgcttccc tgttggtaag gggttaactg tagccctgag gtgggacatt
tgattttaaa aatcagtcat cttggggcgc ttaggttaga ggaatggtag gcagatgctg
tcactccttg cccctcccct cctccttccc acctggaggg gaaatgaaat ctgacaggta
gaaagagggg agttggggtt ctttttctct ctccctccac cagcatcact ctctgcctct
ccctcaaaaa tacgttcctg ggtcaggata tatgttgact ccctagagag ctctggagtc
aacctcctgg ccttcctcca ccctcactct tggccttttc ctgcccccat ttcctctacc
tgtggggcat ggagccacga gcctttgtgt gacggtttgc tttctctctc ctgtctttag
gtgcatggct gcctcctaat cccatagtcc agaggaggca tccctaggac tgcgggcaag
ggagccgcaa gcccagggca gccttgaacc gtcccctggc ctgccctccg gtgggggcca
ggatgctgaa gaagcagtct gcagggcttg tgctgtgggg cgctatcctc tttgtggcct

ggaatgccct gctgctcctc ttcttctgga cgcgcccagc acctggcagg ccaccctcag
tcagcgctct cgatggcgac cccgccagcc tcacccggga agtgattcgc ctggcccaag
acgccgaggt ggagctggag cgcaggcgtg ggctgctgca gcagatcggg gatgccctgt
cgagccagcg ggggagggtg cccaccgcgg cccctcccgc ccagccgcgt gtgcctgtga
ccccccgcgc ggcggtgatt cccatcctgg tcatcgcctg tgaccgcagc actgttcggc
gctgcctgga caagctgctg cattatcggc cctcggctga gctcttcccc atcatcgtta
gccaggactg cgggcacgag gagacggccc aggccatcgc ctcctacggc agcgcggtca
cgcacatccg gcagcccgac ctgagcagca ttgcggtgcc gccggaccac cgcaagttcc
agggctacta caagatcgcg cgccactacc gctgggcgct gggccaggtc ttccggcagt
ttcgcttccc cgcggccgtg gtggtggagg atgacctgga ggtggccccg gacttcttcg
agtactttcg ggccacctat ccgctgctga aggccgaccc ctccctgtgg tgcgtctcgg
cctggaatga caacggcaag gagcagatgg tggacgccag caggcctgag ctgctctacc
gcaccgactt tttccctggc ctgggctggc tgctgttggc cgagctctgg gctgagctgg
agcccaagtg gccaaaggcc ttctgggacg actggatgcg gcggccggag cagcggcagg
ggcgggcctg catacgccct gagatctcaa gaacgatgac ctttggccgc aagggtgtga
cgcacgggca gttctttgac cagcacctca agtttatcaa gctgaaccag cagtttgtgc
acttcaccca gctggacctg tcttacctgc agcgggaggc ctatgaccga gatttcctcg
cccgcgtcta cggtgctccc cagctgcagg tggagaaagt gaggaccaat gaccggaagg
agctggggga ggtgcgggtg cagtatacgg ggagggacag cttcaaggct ttcgccaagg
ctctgggtgt tatggatgac cttaagtcgg gggttccgag agctggctac cggggtattg
tcaccttcca gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct
atgatcctag ctggaattag cacctgcctg tccttcctgg gccccttctt gccacatcat
gagctgaggt gaccacagtc cccaggctgc atcggcctgc ctgtgtttcc ctcttaggtg
catttatctt tttgattttt ccgagtggca tttaagtgca caaatgataa caagaggatt
attctcccgt tctcaaggga gtcagatcag gggaactatt ctagggtatg ttgcggggta
ttaagcagga aaacactgtg tggtgggggg cactgggctt gttggggcca caaatgtcca
cgtcctgagc tttctcctgg agcatgtgca gagagtttgg caacgttcgc tctcttgacc
agaccccttc tccctgactg gctcttccag ccaggcacga gccctccttc tatacctgct
ccccttccca gtggggactg agttatggga gaaggggaca tatttgtggc caaaatgata
ctaaccaaag gggcttcctt gtcagggcct ggtggagttg gtgggtcatc ggggctcact
gcctcctgcc cttctctcct gtctgacccc cacttagccc ttctctcctt gcagcctagc
agtttatagt tctgagatgg aaagttgaag ggggcaagca agacctctcc tcagcccatg
cccagctgtc aggagagagg tgcagggagg aaggccttgt gctgggacaa cctctctctt
gccttacctt cagagaggac tatgccctga ccctccttt ctgaaaatca gtgccctccc
tgttgctcta ggaggctcct gctggcttgg tagaagacag aattcgatct gcctgtccct
ttttcccctg gggtttgaca cacaggctcc tctcagcatg aggtggagca gtgaccaggt
ggagcagtga ccaggacgcc tctggcccag tgctgcccag cctccccgcc cgctcccagg
cgccccatgt cctcacaggc caggacgcca tggcggccgg gagcatgcga.

13. A transformed cell, containing a heterologous sequence of DNA encoding a protein having the amino acid sequence of formula I.

14. The transformed cell of Claim 13, wherein said heterologous DNA sequence has the formula III.

15. The transformed cell of Claim 13, wherein said heterologous DNA sequence has the formula IV.

16. A transformed cell, containing a heterologous sequence of DNA encoding a protein having the amino acid sequence of formula II.

17. The transformed cell of Claim 16, wherein said heterologous DNA sequence has the formula V.

18. The transformed cell of Claim 16, wherein said heterologous DNA sequence has the formula VI.

19. A method for preparing a glycoprotein which is a complex or hybrid N-glycan, comprising:

culturing a cell which produces a precursor high-mannose glycoprotein and which contains a heterologous DNA sequence which encodes a protein having the amino acid sequence of formula I.

20. The method of Claim 19, wherein said heterologous DNA sequence has the formula III.

21. The method of Claim 19, wherein said heterologous DNA sequence has the formula IV.

22. A method for preparing a glycoprotein which is a complex or hybrid N-glycan, comprising:

culturing a cell, which produces a precursor high-mannose glycoprotein and which contains a heterologous DNA sequence which encodes a protein having the amino acid sequence of formula II.

23. The method of Claim 22, wherein said heterologous DNA sequence has the formula V.

24. The method of Claim 23, wherein said heterologous DNA sequence has the formula VI.

Peptide 1:

```
1                   10                  20                  30
W A L G Q I F H N F N Y P A A V V V E D D L E V A P D F F E Y f q
```

Peptide 2:

```
1                 10
L W A E L E P K W P K a
```

Peptide 3:

```
1                 10
F W D D W M R R P E Q
```

Peptide 4:

```
1
T D F F P e
```

Peptide 5:

```
1                 10
D L S Y L Q Q E A Y D R D F 1
```

Peptide 6:

```
1                 10                  20
L F R G R R V H L A P P O T W D G Y D P S W t
```

Peptide 7:

```
1
L G W L
```

Peptide 8:

```
1
A T Y P L
```

Oligonucleotides:

```
2S:  5'-TGG GCI GAA CTI GAA CCI AAA TGG-3'
                G T     G       G

2A:  5'-CCA TTT IGG TTC IAG TTC IGC CCA-3'
           C       C     A C

3S:  5'-TTT TGG GAT GAT TGG ATG CG-3'
         C       C   C         A

3A:  5'-CG CAT CCA ATC ATC CCA AAA-3'
        T        G   G       G

6S:  5'-CAA ACI TGG GAT GGI TAT GAT CC-3'
         G           C       C   C

6A:  5'-GG ATC ATA ICC ATC CCA IGT TTG-3'
         G   G       G         C
```

FIGURE 1

SUBSTITUTE SHEET

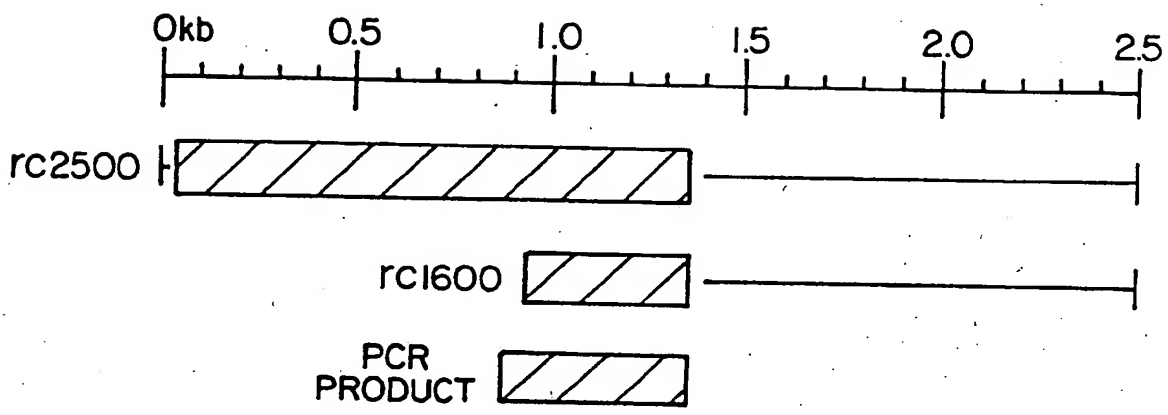


FIGURE 2

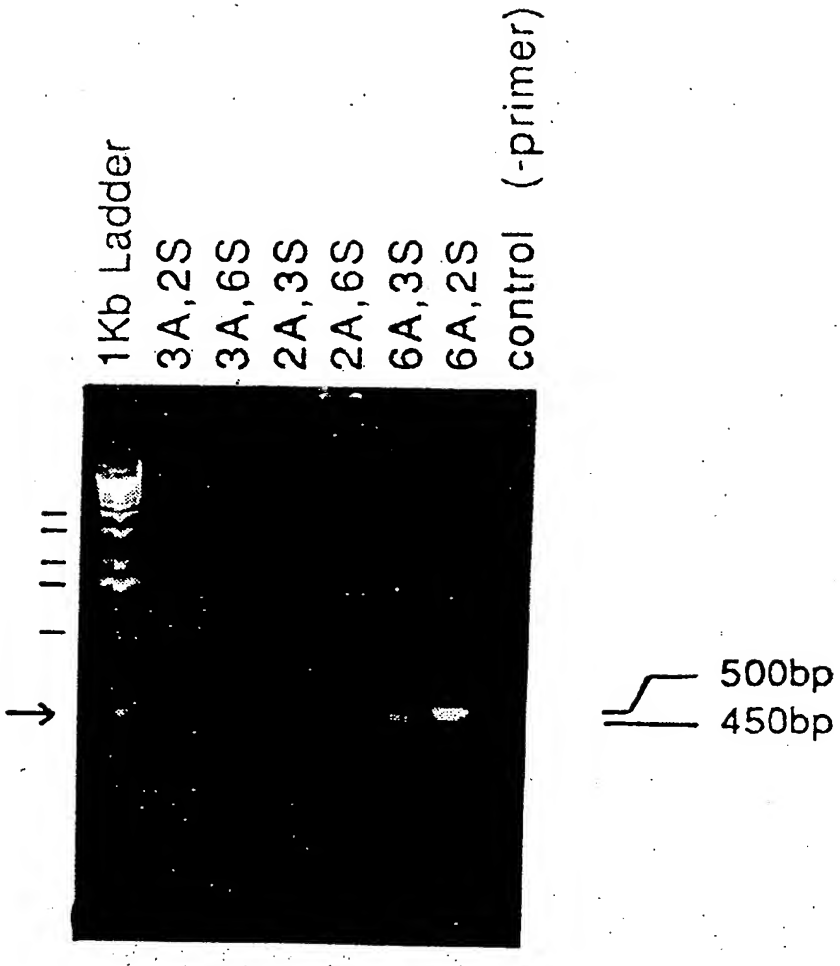


FIGURE 3

```
1 gaattccggc aagtcatacc tttgcctgcc ctcccctgtg ggggccagg

50: atg ctg aag aag cag tct gct ggg ctt gtg ctg tgg ggt gct atc
    MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE

95: ctc ttt gtg gcc tgg aat gcc ctg ctg ctc ctc ttc ttc tgg aca
    LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR

140: cgt cca gtg cct agc agg ctg ccg tca gac aat gct ctc gat gat
     ARG PRO VAL PRO SER ARG LEU PRO SER ASP ASN ALA LEU ASP ASP

185: gac cct gcc agc ctc acc cgt gag gtg atc cgc tta gct cag gat
     ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP

230: gcc gag gta gag ttg gaa cgt cag cgg gga ctg ttg cag cag att
     ALA GLU VAL GLU LEU GLU ARG GLN ARG GLY LEU LEU GLN GLN ILE

275: agg gag cac cat gct ctt tgg agc cag cgg tgg aag gtg cct act
     ARG GLU HIS HIS ALA LEU TRP SER GLN ARG TRP LYS VAL PRO THR

320: gca gcc cct cct gct cag ccg cat gtg cct gtg acc cca ccg cca
     ALA ALA PRO PRO ALA GLN PRO HIS VAL PRO VAL THR PRO PRO PRO

365: gct gtg atc ccc atc ctg gta att gcc tgt gac cgc agc acc gtc
     ALA VAL ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL

410: cgc cgc tgt ttg gac aag cta ctg cat tat cgg cct tca gct gag
     ARG ARG CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU

455: ctg ttc ccc atc att gtc agc cag gac tgt ggg cat gag gag aca
     LEU PHE PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR

500: gcc cag gtc att gct tcc tat ggc agc gca gtc aca cac atc cgg
     ALA GLN VAL ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG

545: caa cct gac ctg agc aac att gct gtg cag ccc gac cac cgc aag
     GLN PRO ASP LEU SER ASN ILE ALA VAL GLN PRO ASP HIS ARG LYS

590: ttc cag ggc tac tac aag atc gca cgg cat tac cgc tgg gca ttg
     PHE GLN GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU

635: ggc caa atc ttc cac aat ttc aac tac cca gca gct gtg gtg gtg
     GLY GLN ILE PHE HIS ASN PHE ASN TYR PRO ALA ALA VAL VAL VAL

680: gaa gat gat ctc gag gtg gca cca gac ttc ttt gag tac ttc cag
     GLU ASP ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE GLN

725: gcc act tac cca ctg ttg aaa gca gac ccc tcc ctc tgg tgt gtg
     ALA THR TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL

770: tct gcc tgg aat gac aat ggc aaa gaa cag atg gta gac tcg agt
     SER ALA TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP SER SER

815: aag cca gag tta ctc tac cgc aca gat ttc ttt cct ggc tta ggc
     LYS PRO GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY
```

FIGURE 4

SUBSTITUTE SHEET

```
860: tgg tta ctg ttg gct gaa ctc tgg gct gaa ctg gag ccc aag tgg
     TRP LEU LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP

905: ccc aaa gcc ttc tgg gat gac tgg atg cgc cgg cct gag cag cga
     PRO LYS ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG

950: aag ggg agg gcc tgt gtg cgt cca gaa atc tca aga aca atg aca
     LYS GLY ARG ALA CYS VAL ARG PRO GLU ILE SER ARG THR MET THR

995: ttt ggc cgg aag ggt gtg agc cat ggg cag ttc ttt gac cag cat
     PHE GLY ARG LYS GLY VAL SER HIS GLY GLN PHE PHE ASP GLN HIS

1040: ctc aag ttc atc aag ctg aac cag cag ttt gta ccc ttc acc cag
      LEU LYS PHE ILE LYS LEU ASN GLN GLN PHE VAL PRO PHE THR GLN

1085: ctg gac ctg tcg tac ctt cag cag gag gcc tat gac cgg gat ttc
      LEU ASP LEU SER TYR LEU GLN GLN GLU ALA TYR ASP ARG ASP PHE

1130: ctt gct cgt gtt tat ggt gct ccc cag tta cag gtg gag aaa gtg
      LEU ALA ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL

1175: agg acc aat gac cgg aag gag cta gga gag gtg cgc gta cag tac
      ARG THR ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR

1220: aca ggc agg gac agc ttc aag gct ttc gcc aag gcc ctg ggt gtc
      THR GLY ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL

1265: atg gat gac ctc aaa tca ggt gta ccc agg gct gga tac cgg ggc
      MET ASP ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY

1310: att gtc acc ttc tta ttc cgg ggc cgc cgt gtc cac ctg gcg ccc
      ILE VAL THR PHE LEU PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO

1355: cct cag act tgg gat ggc tat gat cct agt tgg act
      PRO GLN THR TRP ASP GLY TYR ASP PRO SER TRP THR

1391  taacagctcc tgcctgtccc ttctgggctc cttccttgca atttcatgat ctaagatggg
1451  accgtagtcc ctgggctgca ttgtcttttc tgtctttccc tcttgggtcc attttttttt
1511  ttttcttttt tgagtggcat ttgaatacac agatgacaag gtgagggttc ttttgttaaa
1571  ggagttagat cagggaaagc attctgctgt ctgttgggta tcaagcagca aaccactgtg
1631  tgatagggga agaatgggct ttttggggcc agaaatatcc atgttctgag ttttttcttt
1691  aggtcatctg cagaggagtt ggcaacttta gctttcttaa ccaggccttt tcttttctgac
1751  ctgagagcca gggcatgaga cttcttgttc atgctccttt ttaccttccc ctaataaggg
1811  tctgggctac aggagaagtg aacatattgt ggccagaata atactaacca gaggggcctc
1871  attgtcagag tctaggtgca gttattgggt tgtcagagtt aatgccttct gttcttcttt
1931  ccttattcct gacttctgtc agctcttctt tctttgcagc ctagcaattt ttggttctaa
1991  gatgaaaaat gaagaggaaa agaaatattc gcacccagct attgggagaa aggtagtggg
2051  aaaaaaactt cattgtacca cttcaaagag acactcttga cctcttcctt tctaaaaatt
2111  agtccccctcc ctgttgcttc aggagaatgc tgtgctggtc agttctgtgt gatccttctt
2171  ccctgagttt tatacacagg ctcctcccta aggctgtggc ttctggtggc cctcctgaca
2231  taagttacag tggccaagac caggacaact ccggccatga gctaagtcct gcctaccttc
2291  tccaaaacat tcccatgtcc tcacaggcta ggatgcagat gttggttgga gaggaatttg
2351  tgtgtgtgtg tgtgtgtgtg tgtgttttct tgcctgacct cagtttcatg gatgaaaagt
2411  ggaagctaca gaattatttt caaaaataaa ggctgaattg tctgaaaaaa aaaaaaaaaa
2471  aaaaaaccgg aattc
```

FIGURE 4 (continued)

SUBSTITUTE SHEET

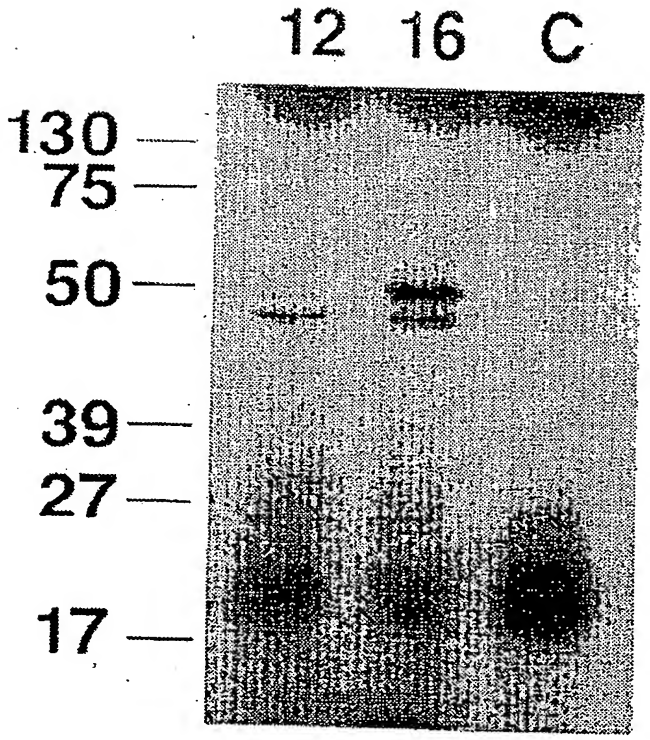


FIGURE 5

```
   1 aagttttgaa tgtttaagtt tatttaagtt tatttctaaa tatttttcta tttctctggc
  61 ttttgtaagt agggttttct catccatgtt ttcttctcat gagttatttg tggatatgaa
 121 ggctatccat tagtatatgt tgatttttat attacacttc cttgctcagt tcattattga
 181 ttctttttga gttttccagg catattctca caagtaaaga taatagaaat agtttgcttc
 241 ctttccactt ctgctttgaa tttttttttc ttggttcatt tgcattggct gcttcctcca
 301 gcaaaatgtt aaataaccct ggagatgatg ggcaacttcg ttttgctcct gacattcgtg
 361 gggtgcctct ggtgcttccc tgttggtaag gggttaactg tagccctgag gtgggacatt
 421 tgattttaaa aatcagtcat cttggggcgc ttaggttaga ggaatggtag gcagatgctg
 481 tcactccttg ccctctcccc cctccttccc acctggaggg gaaatgaaat ctgacaggta
 541 gaaagagggg agttggggtt ctttttctct ctccctccac cagcatcact ctctgcctct
 601 ccctcaaaaa tacgttcctg ggtcaggata tatgttgact ccctagagag ctctggagtc
 661 aacctcctgg ccttcctcca ccctcactct tggccttttc ctgcccccat ttcctctacc
 721 tgtggggcat ggagccacga gcctttgtgt gacggtttgc tttctctctc ctgtctttag
 781 gtgcatggct gcctcctaat cccatagtcc agaggaggca tccctaggac tgcgggcaag
 841 ggagccgcaa gcccagggca gccttgaacc gtcccctggc ctgccctccg gtgggggcca
 901 ggatgctgaa gaagcagtct gcagggcttg tgctgtgggg cgctatcctc tttgtggcct
 961 ggaatgccct gctgctcctc ttctttctgga cgcgcccagc acctggcagg ccaccctcag
1021 tcagcgctct cgatggcgac cccgccagcc tcacccggga agtgattcgc ctggcccaag
1081 acgccgaggt ggagctggag cgcaggcgtg ggctgctgca gcagatcggg gatgccctgt
1141 cgagccagcg ggggagggtg cccaccgcgg ccctccccgc ccagccgcgt gtgcctgtga
1201 ccccgcgccc ggcggtgatt cccatcctgg tcatcgcctg tgaccgcagc actgttcggc
1261 gctgcctgga caagctgctg cattatcggc cctcggctga gctcttcccc atcatcgtta
1321 gccaggactg cgggcacgag gagacggccc aggccatcgc ctcctacggc agcgcggtca
1381 cgcacatccg gcagcccgac ctgagcagca ttgcggtgcc gccggaccac cgcaagttcc
1441 agggctacta caagatcgcg cgccactacc gctgggcgct gggccaggtc ttccggcagt
1501 ttcgcttccc cgcggccgtg gtggtggagg atgacctgga ggtggccccg gacttcttcg
1561 agtactttcg ggccacctat ccgctgctga aggccgaccc ctccctgtgg tgcgtctcgg
1621 cctggaatga caacggcaag gagcagatgg tggacgccag caggcctgag ctgctctacc
1681 gcaccgactt tttccctggc ctgggctggc tgctgttggc cgagctctgg gctgagctgg
1741 agcccaagtg gccaaaggcc ttctgggacg actggatgcg gcggccggag cagcggcagg
1801 ggcgggcctg catacgccct gagatctcaa gaacgatgac ctttggccgc aagggtgtga
1861 cgcacgggca gttctttgac cagcacctca agtttatcaa gctgaaccag cagtttgtgc
1921 acttcaccca gctggacctg tcttacctgc agcgggaggc ctatgaccga gatttcctcg
1981 cccgcgtcta cggtgctccc cagctgcagg tggagaaagt gaggaccaat gaccggaagg
2041 agctggggga ggtgcgggtg cagtatacgg ggagggacag cttcaaggct ttcgccaagg
2101 ctctgggtgt tatggatgac cttaagtcgg gggttccgag agctggctac cggggtattg
2161 tcaccttcca gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct
2221 atgatcctag ctggaattag cacctgcctg tccttcctgg gccccttctt gccacatcat
2281 gagctgaggt gaccacagtc cccaggctgc atcggcctgc ctgtgtttcc ctcttaggtg
2341 catttatctt tttgattttt ccgagtggca tttaagtgca caaatgataa caagaggatt
2401 attctcccgt tctcaaggga gtcagatcag gggaactatt ctagggtatg ttgcggggta
2461 ttaagcagga aaacactgtg tggtgggggg cactgggctt gttggggcca caaatgtcca
2521 cgtcctgagc tttctcctgg agcatgtgca gagagtttgg caacgttcgc tctcttgacc
2581 agaccccttc tccctgactg gctcttccag ccaggcacga gccctccttc tatacctgct
2641 ccccttccca gtggggactg agttatggga gaaggggaca tatttgtggc caaaatgata
2701 ctaaccaaag gggcttcctt gtcagggcct ggtggagttg gtgggtcatc ggggctcact
2761 gcctcctgcc cttctctcct gtctgacccc cacttagccc ttctctcctt gcagcctagc
2821 agtttatagt tctgagatgg aaagttgaag ggggcaagca agacctctcc tcagcccatg
2881 cccagctgtc aggagagagg tgcagggagg aaggccttgt gctgggacaa cctctctctt
2941 gccttacctt cagagaggac tatgccctga cccctccttt ctgaaaatca gtgccctccc
3001 tgttgctcta ggaggctcct gctggcttgg tagaagacag aattcgatct gcctgtccct
3061 ttttcccctg gggtttgaca cacaggctcc tctcagcatg aggtggagca gtgaccaggt
3121 ggagcagtga ccaggacgcc tctggcccag tgctgcccag cctccccgcc cgctcccagg
3181 cgccccatgt cctcacaggc caggacgcca tggcggccgg gagcatgcga
```

FIGURE 6

SUBSTITUTE SHEET

```
  1:  MET LEU LYS LYS GLN SER ALA GLY LEU VAL LEU TRP GLY ALA ILE
 16:  LEU PHE VAL ALA TRP ASN ALA LEU LEU LEU LEU PHE PHE TRP THR
 31:  ARG PRO ALA PRO GLY ARG PRO PRO SER VAL SER ALA LEU ASP GLY
 46:  ASP PRO ALA SER LEU THR ARG GLU VAL ILE ARG LEU ALA GLN ASP
 61:  ALA GLU VAL GLU LEU GLU ARG ARG ARG GLY LEU LEU GLN GLN ILE
 76:  GLY ASP ALA LEU SER SER GLN ARG GLY ARG VAL PRO THR ALA ALA
 91:  PRO PRO ALA GLN PRO ARG VAL PRO VAL THR PRO ALA PRO ALA VAL
106:  ILE PRO ILE LEU VAL ILE ALA CYS ASP ARG SER THR VAL ARG ARG
121:  CYS LEU ASP LYS LEU LEU HIS TYR ARG PRO SER ALA GLU LEU PHE
136:  PRO ILE ILE VAL SER GLN ASP CYS GLY HIS GLU GLU THR ALA GLN
151:  ALA ILE ALA SER TYR GLY SER ALA VAL THR HIS ILE ARG GLN PRO
166:  ASP LEU SER SER ILE ALA VAL PRO PRO ASP HIS ARG LYS PHE GLN
181:  GLY TYR TYR LYS ILE ALA ARG HIS TYR ARG TRP ALA LEU GLY GLN
196:  VAL PHE ARG GLN PHE ARG PHE PRO ALA ALA VAL VAL VAL GLU ASP
211:  ASP LEU GLU VAL ALA PRO ASP PHE PHE GLU TYR PHE ARG ALA THR
226:  TYR PRO LEU LEU LYS ALA ASP PRO SER LEU TRP CYS VAL SER ALA
241:  TRP ASN ASP ASN GLY LYS GLU GLN MET VAL ASP ALA SER ARG PRO
256:  GLU LEU LEU TYR ARG THR ASP PHE PHE PRO GLY LEU GLY TRP LEU
271:  LEU LEU ALA GLU LEU TRP ALA GLU LEU GLU PRO LYS TRP PRO LYS
286:  ALA PHE TRP ASP ASP TRP MET ARG ARG PRO GLU GLN ARG GLN GLY
301:  ARG ALA CYS ILE ARG PRO GLU ILE SER ARG THR MET THR PHE GLY
316:  ARG LYS GLY VAL THR HIS GLY GLN PHE PHE ASP GLN HIS LEU LYS
331:  PHE ILE LYS LEU ASN GLN GLN PHE VAL HIS PHE THR GLN LEU ASP
346:  LEU SER TYR LEU GLN ARG GLU ALA TYR ASP ARG ASP PHE LEU ALA
361:  ARG VAL TYR GLY ALA PRO GLN LEU GLN VAL GLU LYS VAL ARG THR
376:  ASN ASP ARG LYS GLU LEU GLY GLU VAL ARG VAL GLN TYR THR GLY
391:  ARG ASP SER PHE LYS ALA PHE ALA LYS ALA LEU GLY VAL MET ASP
406:  ASP LEU LYS SER GLY VAL PRO ARG ALA GLY TYR ARG GLY ILE VAL
421:  THR PHE GLN PHE ARG GLY ARG ARG VAL HIS LEU ALA PRO PRO PRO
436:  THR TRP GLU GLY TYR ASP PRO SER TRP ASN***
```

FIGURE 7

START

```
880: c c tgc cct ccg gtg ggg gcc agg|atg ctg aag aag cag tct gca
  3: . . CYS PRO PRO VAL GLY ALA ARG↓MET LEU LYS LYS GLN SER ALA

924: ggg ctt gtg ctg tgg ggc gct atc ctc ttt gtg gcc tgg aat gcc
  3: GLY LEU VAL LEU TRP GLY ALA ILE LEU PHE VAL ALA TRP ASN ALA

969: ctg ctg ctc ctc ttc ttc tgg acg cgc cca gca cct ggc agg cca
  3: LEU LEU LEU LEU PHE PHE TRP THR ARG PRO ALA PRO GLY ARG PRO

1014: ccc tca gtc agc gct ctc gat ggc gac ccc gcc agc ctc acc cgg
   3: PRO SER VAL SER ALA LEU ASP GLY ASP PRO ALA SER LEU THR ARG

1059: gaa gtg att cgc ctg gcc caa gac gcc gag gtg gag ctg gag cgc
   3: GLU VAL ILE ARG LEU ALA GLN ASP ALA GLU VAL GLU LEU GLU ARG

1104: agg cgt ggg ctg ctg cag cag atc ggg gat gcc ctg tcg agc cag
   3: ARG ARG GLY LEU LEU GLN GLN ILE GLY ASP ALA LEU SER SER GLN

1149: cgg ggg agg gtg ccc acc gcg gcc cct ccc gcc cag ccg cgt gtg
   3: ARG GLY ARG VAL PRO THR ALA ALA PRO PRO ALA GLN PRO ARG VAL

1194: cct gtg acc ccc gcg ccg gcg gtg att ccc atc ctg gtc atc gcc
   3: PRO VAL THR PRO ALA PRO ALA VAL ILE PRO ILE LEU VAL ILE ALA

1239: tgt gac cgc agc act gtt cgg cgc tgc ctg gac aag ctg ctg cat
   3: CYS ASP ARG SER THR VAL ARG ARG CYS LEU ASP LYS LEU LEU HIS

1284: tat cgg ccc tcg gct gag ctc ttc ccc atc atc gtt agc cag gac
   3: TYR ARG PRO SER ALA GLU LEU PHE PRO ILE ILE VAL SER GLN ASP

1329: tgc ggg cac gag gag acg gcc cag gcc atc gcc tcc tac ggc agc
   3: CYS GLY HIS GLU GLU THR ALA GLN ALA ILE ALA SER TYR GLY SER

1374: gcg gtc acg cac atc cgg cag ccc gac ctg agc agc att gcg gtg
   3: ALA VAL THR HIS ILE ARG GLN PRO ASP LEU SER SER ILE ALA VAL

1419: ccg ccg gac cac cgc aag ttc cag ggc tac tac aag atc gcg cgc
   3: PRO PRO ASP HIS ARG LYS PHE GLN GLY TYR TYR LYS ILE ALA ARG

1464: cac tac cgc tgg gcg ctg ggc cag gtc ttc cgg cag ttt cgc ttc
   3: HIS TYR ARG TRP ALA LEU GLY GLN VAL PHE ARG GLN PHE ARG PHE

1509: ccc gcg gcc gtg gtg gtg gag gat gac ctg gag gtg gcc ccg gac
   3: PRO ALA ALA VAL VAL VAL GLU ASP ASP LEU GLU VAL ALA PRO ASP

1554: ttc ttc gag tac ttt cgg gcc acc tat ccg ctg ctg aag gcc gac
   3: PHE PHE GLU TYR PHE ARG ALA THR TYR PRO LEU LEU LYS ALA ASP

1599: ccc tcc ctg tgg tgc gtc tcg gcc tgg aat gac aac ggc aag gag
   3: PRO SER LEU TRP CYS VAL SER ALA TRP ASN ASP ASN GLY LYS GLU

1644: cag atg gtg gac gcc agc agg cct gag ctg ctc tac cgc acc gac
   3: GLN MET VAL ASP ALA SER ARG PRO GLU LEU LEU TYR ARG THR ASP
```

FIGURE 8

```
1689: ttt ttc cct ggc ctg ggc tgg ctg ctg ttg gcc gag ctc tgg gct
   3: PHE PHE PRO GLY LEU GLY TRP LEU LEU LEU ALA GLU LEU TRP ALA

1734: gag ctg gag ccc aag tgg cca aag gcc ttc tgg gac gac tgg atg
   3: GLU LEU GLU PRO LYS TRP PRO LYS ALA PHE TRP ASP ASP TRP MET

1779: cgg cgg ccg gag cag cgg cag ggg cgg gcc tgc ata cgc cct gag
   3: ARG ARG PRO GLU GLN ARG GLN GLY ARG ALA CYS ILE ARG PRO GLU

1824: atc tca aga acg atg acc ttt ggc cgc aag ggt gtg acg cac ggg
   3: ILE SER ARG THR MET THR PHE GLY ARG LYS GLY VAL THR HIS GLY

1869: cag ttc ttt gac cag cac ctc aag ttt atc aag ctg aac cag cag
   3: GLN PHE PHE ASP GLN HIS LEU LYS PHE ILE LYS LEU ASN GLN GLN

1914: ttt gtg cac ttc acc cag ctg gac ctg tct tac ctg cag cgg gag
   3: PHE VAL HIS PHE THR GLN LEU ASP LEU SER TYR LEU GLN ARG GLU

1959: gcc tat gac cga gat ttc ctc gcc cgc gtc tac ggt gct ccc cag
   3: ALA TYR ASP ARG ASP PHE LEU ALA ARG VAL TYR GLY ALA PRO GLN

2004: ctg cag gtg gag aaa gtg agg acc aat gac cgg aag gag ctg ggg
   3: LEU GLN VAL GLU LYS VAL ARG THR ASN ASP ARG LYS GLU LEU GLY

2049: gag gtg cgg gtg cag tat acg ggg agg gac agc ttc aag gct ttc
   3: GLU VAL ARG VAL GLN TYR THR GLY ARG ASP SER PHE LYS ALA PHE

2094: gcc aag gct ctg ggt gtt atg gat gac ctt aag tcg ggg gtt ccg
   3: ALA LYS ALA LEU GLY VAL MET ASP ASP LEU LYS SER GLY VAL PRO

2139: aga gct ggc tac cgg ggt att gtc acc ttc cag ttc cgg ggc cgc
   3: ARG ALA GLY TYR ARG GLY ILE VAL THR PHE GLN PHE ARG GLY ARG

2184: cgt gtc cac ctg gcg ccc cca ccg acg tgg gag ggc tat gat cct
   3: ARG VAL HIS LEU ALA PRO PRO PRO THR TRP GLU GLY TYR ASP PRO
                 STOP
2229: agc tgg aat|tag cac ctg cct g
   3: SER TRP ASN↓*** HIS LEU PRO .
```

FIGURE 8 (continued)

SUBSTITUTE SHEET

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/54, 9/10

A3

(11) International Publication Number: WO 92/09694

(43) International Publication Date: 11 June 1992 (11.06.92)

(21) International Application Number: PCT/CA91/00417

(22) International Filing Date: 29 November 1991 (29.11.91)

(30) Priority data:
620,098 30 November 1990 (30.11.90) US

(71) Applicant: HSC RESEARCH AND DEVELOPMENT LIMITED PARTNERSHIP [CA/CA]; 88 Elm Street, Toronto, Ontario M5G 1X8 (CA).

(72) Inventors: SCHACHTER, Harry ; 5 Menin Road, Toronto, Ontario M6C 3J1 (CA). SARKAR, Mohan ; 77 Elm Street, Apartment #405, Toronto, Ontario M5G 1H4 (CA).

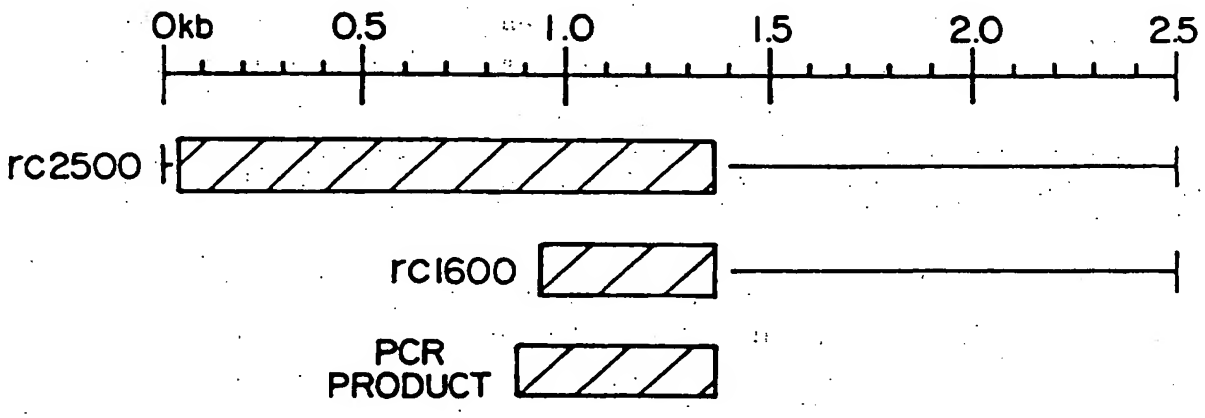
(74) Agent: D'IORIO, Hélène; Gowling, Strathy & Henderson, Suite 2600, 160 Elgin Street, Ottawa, Ontario K1N 8S3 (CA).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, PL, RO, SD, SE, SE (European patent), SN (OAPI patent), SU+,TD (OAPI patent), TG (OAPI patent).

Published
With international search report.

(88) Date of publication of the international search report:
10 October 1996 (10.10.96)

(54) Title: CLONING OF UDP-N-ACETYLGLUCOSAMINE:α-3-D-MANNOSIDE β-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I



(57) Abstract

The genes encoding rabbit and human GnT I have been cloned.

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

INTERNATIONAL SEARCH REPORT

PCT/CA 91/00417

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl.5 C 12 N 15/54 C 12 N 9/10

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl.5	C 12 N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X	Glycoconjugate Journal, vol. 7, no. 5, 10 October 1990, (Lund, SE), E. HULL et al.: "Isolation of 13 and 15 kilobase human genomic DNA clones containing the gene for UDP-N-acetylglucosamine:alpha-3-D-mannoside beta-1,2-N-acetylglucosaminyltransferase I", page 468, abstract no. 85, see the whole document ---	1-3,7-9 ,13-15
X	Glycoconjugate Journal, vol. 7, no. 5, 10 October 1990, (Lund, SE), M. SARKAR et al.: "Rabbit liver UDP-N-acetylglucosamine:alpha-3-D-mannoside beta-1,2-N-acetylglucosaminyltransferase I: characteriazation of a 2,5 kilobase cDNA clone", page 380, abstract no. 4, see the whole document --- -/-	1-3,7-9 ,13-15

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
18-02-1992	15.04.93
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	T.K. WILLIS

Form PCT/ISA/210 (second sheet) (January 1985)

See notes on accompanying sheet

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
P,X	Proc. Natl. Acad. Sci. USA, vol. 88, no. 1, January 1991, Natl. Acad. Sci., (Washington, DC, US), M. SAKKAR et al.: "Molecular cloning and expression of cDNA encoding the enzyme that controls conversion of high-mannose to hybrid and complex N-glycans: UDP-N-acetylglucosamine:alpha-3-D-mannoside beta-1,2-N-acetylglucosaminyltransferase I", pages 234-238, see figure 4; page 236, left-hand column, line 26 - page 237, right-hand column, line 6 ---	1-3,7-9 ,13-15
P,X	Biochem. Soc. Trans., vol. 19, no. 3, August 1991, Biochemical Society, (London, GB), H. SCHACHTER et al.: "Molecular cloning of human and rabbit UDP-N-acetylglucosamine:alpha-3-D-mannoside beta-1,2-N-acetylglucosaminyltransferase I", pages 645-648, see page 646, left-hand column, line 1 - page 648, right-hand column, line 23 ---	1-3,7-9 ,13-15
Y	J. Biol. Chem., vol. 263, no. 17, 15 June 1988, Am. Soc. Biol. Chem., Inc., (US), Y. NISHIKAWA et al.: "Control of glycoprotein synthesis. Purification and characterization of rabbit liver UDP-N-acetylglucosamine:alpha-3-D-mannoside beta-1,2-N-acetylglucosaminyltransferase I", pages 8270-8281, see table I; abstract; page 8270, right-hand column, lines 25-29 (cited in the application) ---	1-3,7-9 ,13-15, 19-21
Y	J. Biol. Chem., vol. 265, no. 2, 15 January 1990, Am. Soc. Biol. Chem., Inc., (US), F. YAMAMOTO et al.: "Cloning and characterization of DNA complementary to human UDP-GalNAc: Fucalpha1 2Gal alpha1 3GalNAc transferase (histo-blood group A transferase) mRNA", pages 1146-1151, see materials and methods (cited in the application) --- -/-	1-3,7-9 ,13-15, 19-21

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	J. Biol. Chem., vol. 256, no. 2, 25 January 1981, Am. Soc. Biol. Chem., Inc., (US), C.L. OPPENHEIMER et al.: "Purification and characterization of a rabbit liver alphal 3 mannoside betal 2 N-acetylglucosaminyltransferase", pages 799-804, see page 801, left-hand column, line 8 - right-hand column, line 8 (cited in the application)	1-3,7-9 ,13-15, 19-21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA91/00417

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. CLAIMS: 1-3,7-9,13-15,19-21
2. CLAIMS: 4-6,10-12,16-18,22-24

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-3,7-9,13-15,19-21

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.